Title

Molecular and Physiological Evolution of the Chemosensory Systems in a Mustard-Feeding Drosophilid Fly

Author

Suzuki, Hiromu

Publication Date

2024

Peer reviewed|Thesis/dissertation

Molecular and Physiological Evolution of the Chemosensory Systems in a Mustard-Feeding Drosophilid Fly

By

Hiromu Suzuki

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Integrative Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Noah Whiteman, Chair
Professor Kristin Scott
Professor Rebecca Tarvin
Professor Diana Bautista

Spring 2024

Abstract

Molecular and Physiological Evolution of the Chemosensory Systems
in a Mustard-Feeding Drosophilid Fly

By

Hiromu Suzuki

Doctor of Philosophy in Integrative Biology

University of California, Berkeley

Professor Noah Whiteman, Chair

The remarkable diversity of animal behaviors have long attracted biologists from different fields. However, the genetic architecture and molecular changes behind the diversification of these behavioral traits have remained elusive due to technical limitations, preventing deeper understanding of evolutionary processes of behaviors. My dissertation work aims to dissect the evolution, genetic architecture, and molecular basis of behavioral traits using a herbivorous drosophilid fly, *Scaptomyza flava*, with a particular emphasis on the evolution of its chemosensory behaviors that are specifically tuned to the specialized chemicals produced by mustard plants in the order Brassicales.

In Chapter 1, I briefly summarized the historical context and recent trends in the study of behavioral evolution in animals. I argue that the evolution of chemosensory behavior in insects provides an excellent model for uncovering the genetic architecture and molecular basis of behavioral change across lineages. I further introduced *S. flava* as a unique system to address such questions that allows us to bridge genomic and functional approaches by leveraging the power of model organisms in genetics, the vinegar fly *Drosophila melanogaster* and the model mustard *Arabidopsis thaliana*.

In Chapter 2, I investigated transcriptional evolution in the chemosensory organs, in the genus *Scaptomyza* with a particular emphasis on six major insect chemosensory gene families. I conducted a comparative bulk RNA-sequencing between *S. flava* and its microbe-feeding relative, *S. pallida,* and searched for genes experiencing larger expression divergence between species as candidates behind chemosensory evolution.

This approach successfully illuminated evolutionary patterns across gene families and narrowed down the candidates that provide insight into specific sensory changes aided the colonization of Brassicales host plants.

In contrast to the exploratory analyses in the previous chapter, Chapter 3 conducted a detailed investigation on a single gene and its products, *TrpA1*, which encodes a chemosensory ion channel detecting electrophilic toxins, including mustard-derived isothiocyanates (ITCs). Again, I took a comparative approach between the mustard-feeding *S. flava* and the microbe-feeding relatives to address how the functional evolution of TRPA1 has been intertwined with the dietary and chemosensory evolution among the drosophilids. I integrated multiple experimental approaches spanning from genomics and genetics to electrophysiology, and found that the reduced chemical sensitivity of the TRPA1 channel and the modified composition of alternatively spliced isoforms were likely two major molecular changes behind the weak avoidance toward ITCs in the *S. flava* lineage. However, I also found that the TRPA1 activity from the distantly-related generalist *D. melanogaster* was even weaker than *S. flava*, despite the strong behavioral aversion to ITCs, suggesting that the evolutionary relationship of the TRPA1 channel function and the gustatory preferences is more complicated than it seems, and requires careful investigations on multiple characteristics of the gene.

To Mika chan

Table of Contents

Acknowledgements

First, I would like to thank my advisor Noah Whiteman for his support in the last seven years during my Ph.D. I cannot even imagine earning a degree without his mentorship. I still remember the day he took me for a walk to Strawberry Creek to show me leaf-mining creatures in the wild when I was struggling in the first year of grad school. His deep knowledge and love for the natural world was always inspirational. His encouragement gave me confidence to pursue my own interests, and helped me to develop perseverance during tough times. He showed me a radically different figure from what I knew as a scientist, and I have learned countless things from our conversations.

I thank my dissertation committee members, Kristin Scott, Diana Bautista, and Becca Tarvin for their critical feedback on my work. Receiving feedback and discussing my work with such giants were a huge privilege for me.

I am grateful to all my labmates in the Whiteman lab who shared their time with mine: Tim O'Connor, Cathy Rushworth, Rebecca Duncan, Kirsten Verster, Nicolas Alexandre, Julianne Peláez, Teru Matsunaga, Jessica Aguilar, Marianna Karageorgi, Susan Bernstein, Diler Haji, Carolina Reisenman, Rebecca Tarnopol, Moe Bakhtiari, Ben Goldman-Huertas, and Sam Chaturvedi. I love how each single person had a unique expertise and personality who taught me a lot and made my Ph.D. days more exciting. I would also like to give a special shoutout to an undergraduate researcher Ashleigh Suzu Takemoto for her substantial support of my research and her effort in fly/plant cares in the past few years. I also thank her for our occasional chats in Japanese.

I thank all my collaborators for their valuable support and intellectual contributions, including Shiegru Saito, Claire Saito, Makoto Tominaga, Takaaki Sokabe, Shoma Saito, Megan Laturney, Varsha Rajshekar, Gary Karpen, Lydia Smith, John V. Lin King, David Julius, Juliana Valencia Lesmes, Oscar Arenas, Kentaro Tanaka, and Aya Takahashi. Some of the collaborators kindly hosted me in their groups for several months and allowed me to learn different lab cultures, which were valuable moments that deepened my knowledge and expanded my perspectives.

I am thankful to Masakado Kawata, Takashi Makino, and Katsuhisa Ozaki, who helped me shape my research interests at the earliest stage of my career, and supported my application to the US graduate programs. I am also grateful to Kenkichi Nakajima Memorial Fellowship and Jasso Long-term Fellowship for providing financial support for

the first several years in my Ph.D. I would not be able to pursue a Ph.D. in the US without their help.

I thank all my family members in Japan, especially my parents, Yumiko and Manabu Suzuki, who nurtured my curiosity for nature in my childhood and have supported me distantly in every possible way over the past thirty years. My skills in scientific outreach have probably improved thanks to them. I also thank my late grandmother, Kyoko, who passed away during the writing of this dissertation. I never followed a career path she always wanted me to pursue – staying with parents and working at a local city office – but she loved me throughout her life and so did I. I hope my work is somewhat enjoyable to her.

Finally, I would like to deliver the biggest gratitude to my wife, Mikana. Your intelligence and eyes for details are always inspirational to me. Your jokes always lift my mood and give me energy to survive everyday life. Your kindness turned some of the toughest times in my Ph.D. into memorable moments. I always love the times when you listen to and ask questions on my geeky talks on bugs, even though you hate bugs. Let's hope we can settle in a permanent job in the foreseeable future. No more moving.

Chapter 1. Introduction: The genetic basis of chemosensory evolution in insects.

Abstract

Behavioral evolution has long gained attention from biologists from several fields, but its genetic architecture has been elusive due to technical limitations in a vast majority of “non-model” organisms. However, recent innovations in genome sequencing and genome editing technologies have opened the door for addressing this question for a wide array of organisms with unique behavioral phenotypes. In this chapter, I briefly summarized recent efforts on the genetic dissection of behavioral variations, with a particular emphasis on chemosensory evolution in insects. I further introduce a mustard-feeding drosophilid fly, *Scaptomyza flava*, as a model system of chemosensory evolution with its unique strengths and challenges.

A new era of behavioral evolution studies

Evolution of behavioral traits of animals – courtship, communication, sociality etc. – have gained massive attention from biologists since Charles Darwin (Ekman 2009). Ecologists and evolutionary biologists since the 20th century have succeeded in revealing ultimate causes (per Niko Tinbergen’s criteria; (Tinbergen 1963)) and evolutionary processes of certain evolutionary traits based on behavioral observations and comparative approaches across lineages. One of the emerging trends in this vein is the application of genomic and genetic tools to understand the molecular basis of behaviors across taxa, which potentially allows to track evolutionary dynamics of behavioral traits at a higher resolution.

Historically, discoveries on the cellular and genetic basis of behavior have been achieved by geneticists and neurobiologists, which successfully built a public notion that a certain fraction of behavioral traits are encoded by genetic factors (e.g., (Konopka and Benzer 1971; Dudai et al. 1976; Osborne et al. 1997; Flint 2003; Bendesky and Bargmann 2011)). However, most of those work were performed on a limited number of genetically tractable so-called “model organisms”, such as mouse (*Mus musculus*), fruit fly (*Drosophila melanogaster*), and nematode (*Caenorhabditis elegans*), which are distantly located to one another on a species tree. Although such studies uncovered basic principles of nervous systems across animals, they were essentially helpless on understanding how behavioral differences between lineages arose during evolution. On the other hand, evolutionary biologists mostly worked on “non-model” organisms with unique behavioral traits where virtually no genetic information or tools were available. These gaps between the fields have prevented the understanding of behavioral evolution at the molecular level for decades (Jourjine and Hoekstra 2021) .

However, recent technological innovations have been changing the game. One of them is the advancement of nucleotide sequencing technology. Two decades after the first report of the whole genome sequence of *Homo sapiens (Lander et al. 2001)*, achieved by a thirteen-year effort solely dependent on Sanger sequencers, it has become extremely easier and cheaper to get a whole draft genome sequenced for any given species, generally within a week (Goodwin et al. 2016). Current standards are high-throughput short-read sequencers (e.g., Illumina NovaSeq/HiSeq series) and long-read sequencers (e.g., Pacific Biosciences, Oxford Nanopore Technologies), combined with bioinformatic genome assembling softwares. Another factor is the development of genome editing tools, including CRISPR/Cas9 (Doudna and Charpentier 2014; Jiang and Doudna 2017). Although other genome editing tools have been developed in the more distant past, such as ZFN or TALEN, the CRISPR/Cas9 system allows genome editing of any genomic sequence of any organisms through a relatively simple experimental protocol. This tool revolutionized the study of a number of fields in biology. Taken together, these technologies have revolutionized the study of behavioral evolution by potentially allowing the genomic scan to identify candidate genes and the modification of the candidate genes (or cells) in virtually any organisms.

Insect chemoreception as a model of behavioral evolution studies

Insects are the most species-rich group in animals, and the diversity and intricacy of behavioral traits across lineages – feeding, mating, migration, and social structures – has been attracting the attention of many biologists. Because of the high dependence on environmental chemical cues and the variation of ecological niches, chemosensory behavior has diverged across lineages thus has been an appealing study target. In this section, I briefly summarize the history of the studies of insect chemosensory evolution from the genetic perspectives, latest achievements, and future directions of the field.

Our knowledge on the relationships between genes and behaviors in insects has been traditionally developed by classic genetic works using *Drosophila melanogaster*. Isolation of behavioral mutants and subsequent genetic clonings successfully identified genes underlying fly behavior, with famous examples like *period* (circadian rhythm; (Konopka and Benzer 1971; Jackson et al. 1986)), *dunce* (associative learning; (Dudai et al. 1976)), *foraging* (feeding activity; (Osborne et al. 1997)), and *satori/fruitless* (courtship; (Ito et al. 1996)). Despite these rich experimental resources, however, genes/molecules underlying chemosensory detection in the flies had been elusive for a long time, because of the lack of mutants with suitable phenotypes. This situation changed when the whole genome of *D. melanogaster* was sequenced (Adams et al. 2000). Prior to this report, vertebrate odorant receptor gene families were identified in the mouse genome, which encoded G-protein coupled receptors (GPCRs) with a seven-transmembrane domain (Buck and Axel 1991). Taking advantage of this

discovery, several teams took a computational approach to search for sequences in the *D. melanogaster* genome encoding similar seven-transmembrane structured proteins, and successfully identified insect olfactory receptor (*Ors*) and gustatory receptor (*Grs*) gene families expressed in chemosensory organs of flies (Clyne et al. 1999; Vosshall et al. 1999; Clyne et al. 2000; Kim et al. 2000; Scott et al. 2001; Robertson et al. 2003). Subsequent work further discovered architectures of fly chemosensation, such as a singular gene expression of *Ors* in olfactory neurons and the presence of olfactory co-receptor, ORCO (Larsson et al. 2004), detection of sugars via GRs (Dahanukar et al. 2007; Jiao et al. 2007; Jiao et al. 2008; Fujii et al. 2015), and another class of chemosensory receptor, ionotropic receptors (IRs) (Benton et al. 2009), taking advantage of the wealth of genetic tools and the binary expression system available in *D. melanogaster*. All these findings on the principles of *D. melanogaster* chemosensory systems have been founded on the whole genome sequence.

What about the evolution of chemosensory systems? Again, genomes paved the way forward. In 2007, whole genome sequences of 11 species in the genus *Drosophila* were reported (Drosophila 12 Genomes Consortium et al. 2007), which subsequently spurred the comparative analyses of genomic components and molecular evolution across fly lineages (Makino and Kawata 2012). In the context of behavioral evolution, *Drosophila sechellia* has been a focus of attention for many biologists. *D. sechellia* is an endemic species in the Seychelles archipelago diverged from *D. simulans* lineage approximately ~0.24 million years ago (Garrigan et al. 2012). Unlike the generalist sister species like *D. simulans* and *D. melanogaster*, *D. sechellia* exclusively feeds and reproduces on the ripe noni fruit, *Morinda citrifolia*, enriched with toxic organic acids and methyl esters that attracts *D. sechellia* but repels other generalist species (Auer et al. 2021). This drastic shift in chemosensory preferences between sister species offers one of the most ideal platforms of comparative analysis of behavioral evolution. Seminal work on this topic were (McBride 2007) and a subsequent paper by (McBride and Arguello 2007), where the authors found rapid loss of olfactory and gustatory receptor genes in *D. sechellia* by thorough genomic scans and comparisons between species (although this gene loss might also be confounded by small effective population size of island species; (Gardiner et al. 2008)). At the same time, (Matsuo et al. 2007) investigated the genetic basis of host preference for noni fruits through hybrid crosses between *D. melanogaster* and *D. sechellia* and behavioral assay, and successfully identified that the locus near two odorant-binding protein genes, *Obp57d* and *Obp57e*, underlie their species-specific preferences to noni-derived odors. This research demonstrated that chemosensory gene families were indeed one of the key drivers of behavioral evolution in *D. sechellia*. Fueled by these findings, genetic basis of chemosensory evolution across drosophilids has been actively studied in comprehensive genus-wide analyses (Sánchez-Gracia et al. 2009; Almeida et al. 2014)

or with a particular focus on ecological specialists like the fruit-feeder *D. suzukii* (Karageorgi et al. 2017; Dweck et al. 2021) or the cactus specialist *D. mojavensis* (Diaz et al. 2018; Khallaf et al. 2020). In parallel, research on *D. sechellia* has also been progressed since then (Shiao et al. 2015; Prieto-Godino et al. 2016; Prieto-Godino et al. 2017; Auer et al. 2020; Álvarez-Ocaña et al. 2023). Most of these studies focused on chemosensory receptors like ORs or IRs and revealed their functional significance on the specialization onto the noni fruits. Notably, Auer et al. (2020) applied CRISPR/Cas9-mediated introduction to precisely swap an olfactory receptor gene in *D. sechellia* genome, *Or22a*, with its orthologous copy from *D. melanogaster*, and demonstrated that the chemical tuning of OR22a underlied species-specific preference for noni fruits. This type of causational, not correlational, analysis of chemosensory evolution will make a new standard for future studies, which will provide us deep understandings of how genes manipulate nervous systems and behavioral output.

Advancement of chemosensory evolution research in *Drosophila* also spurred the investigations on other insect lineages. Within a decade after the first *D. melanogaster* genome report, reference genome sequences from other insect species started to be sequenced, including malaria mosquito (Holt et al. 2002), silkworm moth (Mita et al. 2004), honeybee (Honeybee Genome Sequencing Consortium 2006), and red flour beetle (Richards et al. 2008), to name but a few. Although insect chemosensory genes are generally fast-evolving, they still retain similar protein sequences and structures across lineages, thus they can be easily identified by a series of bioinformatic pipelines. As genome and transcriptome data have accumulated over the years, chemosensory genes have also been identified in multiple different lineages. As a result, comparisons of chemosensory gene sequences, family sizes, or expression profiles have become one of the most prominent research areas in insect behavioral evolution studies (Benton 2015; Robertson 2019).

However, our understanding of chemosensory gene functions – namely, what exact chemical ligands they interact with and how those relationships changed over evolutionary times – has largely fallen behind, primarily because common approaches on chemoreceptor deorphanization are time-consuming and only allows investigating a limited number of receptor/ligand combinations. For example, single-sensillum recordings (SSR) combined with *Drosophila* genetics (i.e., empty neuron system, (Dobritsa et al. 2003; Hallem and Carlson 2006)) provides a powerful *in vivo* tool to isolate a single chemoreceptor and to investigate its chemical responses. SSR can also be combined with the gas chromatography setup (called GC-SSR) allowing the simultaneous characterization of bioactive compounds and their receptors, which successfully deorphanized chemoreceptors in several non-model systems (e.g., (McBride et al. 2014; Brand et al. 2020). However, this approach requires time-consuming fly transgenesis and crosses for several generations, thus is not

suitable for investigating many chemoreceptors at a time (Fleischer et al. 2018). Electrophysiology or calcium imaging using cell-based systems offer faster and more practical options, such as *Xenopus* oocytes or human cultured cell lines, but their throughput is not very high either. On top of that, those cell systems lack molecular agents available in the native neurons or adjacent cells (e.g., OBPs in sensillum lymph) that might modulate receptor functions, therefore it tends to be difficult to interpret the results from these *in vitro* experiments (Fleischer et al. 2018). Gene knockout or knockdown would also be common approaches when the target genes are not amenable to *in vitro* assays, but such experiments are obviously time-consuming and interpretation can be difficult if the phenotypic changes are subtle. Technical innovations on chemoreceptor deorphanization tools, such as a large-scale screening of multiple receptors in a near-native cellular environment, will greatly enhance our understanding of the consequences of chemosensory gene evolution we have just started to recognize.

Scaptomyza flava as a model of behavioral evolution studies

The genus *Scaptomyza* is a lineage within the paraphyletic *Drosophila* genus that recently diverged from Hawaiian *Drosophila* clade (Lapoint et al. 2013). This genus is characterized by the diversity of its dietary ecology: although most species are considered microbe-feeders, several lineages evolved into ecological specialists like flower feeders, mushroom specialists, and predators of spider eggs (Lapoint et al. 2013). One of those extremes is a clade of true herbivores containing 8-10 species, presumably transitioned from ancestral microbe feeders ca. 10-15 million years ago (Lapoint et al. 2013; Katoh et al. 2017; Peláez et al. 2022). This transition to herbivory occurred much more recently compared to the origins of major herbivorous insect clades (e.g., Lepidoptera, many phytophagous Coleoptera), offering an ideal platform to uncover the molecular basis underlying the evolution of herbivory, a key trait behind the massive diversification of arthropods (Wiens et al. 2015).

One of the representative species in the herbivorous *Scaptomyza* clade, called *Scaptomyza flava*, offers a few special benefits in the research of chemosensory evolution. First, *S. flava* is an obligate specialist herbivore of the plants in the mustard family, Brassicaceae, including the genetic model plant *Arabidopsis thaliana* (Whiteman et al. 2011). Female *S. flava* uses its cutting ovipositor to create “feeding punctures” on the surface of host plant leaves to consume plant exudates (Whiteman et al. 2011). Females also occasionally deposit the eggs inside the leaf tissues so its offspring can spend the entire larval stages as a leaf-miner, which confers not only food sources but protection from desiccation and parasitoid attacks in nature (Whiteman et al. 2011). Such high dependence on the mustard host plants implies that both larval and adult *S. flava* frequently encounter defensive chemicals of the mustards. Because

phytochemistry of the Brassicales, and most notably that of *A. thaliana*, has been intensively studied (Brown et al. 2003), it is relatively easier to narrow down ecologically relevant compounds for *S. flava* with which their chemosensory proteins might interact, compared to herbivores on the non-model plants or microbe-feeders. *Arabidopsis* also offers numerous mutant lines including the ones with defects on phytochemical productions, allowing the precise manipulation on the host phenotype to investigate the behavioral responses of *S. flava* (Whiteman et al. 2011), which is not possible in other specialized drosophilids like *D. sechellia* or *D. suzukii*. Second, being closely related to *D. melanogaster* is another major advantage, primarily due to the vast array of knowledge on its chemosensory behavior. For example, expression profiles and ligands of most *D. melanogaster* chemoreceptors (especially for ORs; (Münch and Galizia 2016)) have been previously identified, and such information can be used to extrapolate the ligands of unknown chemoreceptors in *S. flava* as a starting point. Most behavioral experiments developed from *D. melanogaster* can be applied in *S. flava* and other sister drosophilids, which is useful for comparative analyses. These characteristics make *S. flava* a unique system to address evolutionary changes in the chemosensory systems accompanying the drastic shift to herbivory. Indeed, some of those changes in *S. flava* have been uncovered in the previous studies, such as the loss yeast-sensing olfactory receptors (Goldman-Huertas et al. 2015) and the gain of olfactory receptors detecting isothiocyanates, defensive compounds specific to the Brassicales (Matsunaga et al. 2022).

There are also several challenges in using *S. flava* as a model system. The biggest one is the difficulty of applying genome editing tools. Common approaches for CRISPR-mediated transgenesis in drosophilids require microinjection into early-stage embryos. However, due to the peculiar reproductive ecology of *S. flava* laying eggs deep inside the leaf tissue (Whiteman et al. 2011), it is extremely difficult to collect a sufficient number of eggs for microinjection. Transgenesis through direct adult injection of Cas9 ribonuclease has been developed and applied in multiple insect species (Chaverra-Rodriguez et al. 2018; Shirai et al. 2023), but it was also reported that this method might not be compatible with drosophilids (Shirai et al. 2023). Moreover, although *S. flava* can be maintained in the laboratory using *A. thaliana* as a host plant, handling *S. flava* colony takes much longer than the maintenance of regular drosophilids using cornmeal vials, which makes it potentially challenging to isolate and maintain transgenic individuals of *S. flava*. Despite their unique adaptation and potentials, these issues still persist and prevent *S. flava* to be a truly “genetically tractable” system for behavioral evolution like *D. sechellia*. Efforts are still warranted to enable transgenesis in *S. flava*, and once the issue is resolved, this system will provide far richer opportunities for understanding the molecular basis of chemosensory evolution, or that of plant-insect interactions in general. Alternatively, non-transgenic approaches such as

comprehensive screenings of *S. flava* chemoreceptors using SSR or comparative connectomics across species will shed light on the molecular and cellular underpinnings of behavior evolution from different angles, which would complement the genetic approaches.

Chapter 2. Comparative transcriptome analysis of chemosensory organs in the genus *Scaptomyza*.

Experiments and analyses in this chapter were also contributed by the following co-authors: Teruyuki Matsunaga, Julianne N. Pelaez, Benjamin Goldman-Huertas, and Noah K. Whiteman (University of California, Berkeley).

Abstract

Insect chemoreception provides a promising platform to investigate the genetic underpinnings behind behavioral evolution. *Scaptomyza flava* is a recently evolved herbivore nested in the *Drosophila* lineage that has been gaining attention as an emerging model system to study this topic. Here, we focused on understanding how gene expression evolution is associated with the evolution of the chemosensory system in this lineage. Accordingly, in this chapter, we conducted a series of interspecific transcriptome analyses between the herbivore *S. flava* and its related microbe-feeder, *S. pallida*, with a particular emphasis on the major six gene families. Our analyses highlighted general evolutionary patterns across gene families and major expression divergence of *S. flava* in two gene families encoding odorant-binding protein (*Obp*) and pickpocket (*Ppk*) proteins. This work provided the full profiles of chemosensory transcriptomes of the specialist herbivore, and will be a foundation for functional experiments on the evolution of the chemosensory system in the future.

Introduction

Behavioral traits are often quantitative traits underpinned by complex genetic architectures (Anholt and Mackay 2004). Given the complexity of the genetic architecture of a behavioral trait in a single species, it is even more challenging to understand how genetic variation gives rise to new behavioral traits over deeper evolutionary timescales, because addressing such questions requires detailed investigation at the molecular level in multiple species (Bendesky and Bargmann 2011). Insect chemosensation plays an essential part of innate behaviors important for fitness such as courtship and feeding, and is therefore an excellent model to study behavioral evolution. Another reason is that evolutionary shifts in such behaviors can often be attributed to a few molecules in the peripheral sensory neurons.

A mustard-feeding drosophilid fly, *Scaptomyza flava*, is a promising model species to study the molecular basis of chemosensory evolution. Unlike their drosophilid relatives that are mostly microbe-feeders, both larvae and female adults of *S. flava* actively feed on the leaves of their host plants in the order Brassicales, including the genetic model plant *Arabidopsis thaliana*, which produce secondary metabolites

chemically distinct from microbial diets (Whiteman et al. 2011; Aguilar et al. 2024). This drastic dietary shift evolved ca. 10-15 million years ago (MYA) (Lapoint et al. 2013; Peláez et al. 2022), making this lineage suitable for a comparative genomics approach that aims to narrow down the candidate genes underlying the colonization to the mustards. A well-studied catalog of phytochemicals from the genetic model plant *Arabidopsis thaliana* allows straightforward deorphanization of ecologically-relevant chemoreceptors in *S. flava*. Although transgenic tools are not available in *S. flava* yet, its basic biology is well characterized and the rich functional knowledge in the chemosensory system of *D. melanogaster* (diverged ca. 60-70 MYA) allows us to infer the phenotypic consequences of particular genetic changes along the lineage leading to *S. flava*. These characteristics of *S. flava* provide an ideal model system for the research of chemosensory evolution in insects.

Integral to insect chemosensation are gene families that encode receptor proteins or related accessory proteins (Montell 2021). Olfactory receptor (ORs) and gustatory receptor (GRs) are the two core families that are responsible for the detection of a wide array of volatile and soluble compounds in insects. While these two receptors share the structural property with vertebrate chemoreceptors which are G-protein coupled receptors, insect ORs and GRs compose of a distinct clade of ligand-gated ion channels that can directly transduce chemical stimuli. ORs in sensory neurons form a heteromeric ion channel that consists of two protein subunits: One divergent unit that determines the ligand specificity of the receptor, and another conserved unit called olfactory co-receptor (ORCO), which is required for trafficking to neural dendrites and always accompanied to a tuning subunit (Larsson et al. 2004). This functional modularity of ORs allows for the comprehensive characterization of ligands for each single tuning OR in *D. melanogaster (Münch and Galizia 2016)* and in other species (Carey et al. 2010; de Fouchier et al. 2017). GRs are mainly in charge of detecting sugars and nutrients as appetitive or “sweet” tastes, and environmental (mostly plant-derived) non-volatile toxins as aversive or “bitter” tastes (Montell 2009). Sweet and bitter compounds are detected by distinct subfamilies within the GRs, in which bitter receptor gene repertoires substantially vary between species, likely reflecting the diversity of toxins present in different niches (Robertson 2019). Insects also possess other classes of chemosensory ion channels, such as ionotropic receptor (IRs), pickpocket (PPKs), and transient receptor potential (TRPs) families. IRs were derived from another family of ionotropic glutamate receptors (iGluR) mediating synaptic transmission. IRs were originally discovered as another class of odorant receptors for acids and amines, but later found to be also involved in contact chemoreception of amino acids and salts (Benton et al. 2009; Croset et al. 2010; McDowell et al. 2022). PPKs are part of the larger Degenerin/Epithelial Sodium Channel (DEG/ENaC) family found across animals. Although the exact functions of most insect PPKs remain obscure, some members of

PPKs are known to be involved in chemoreception in *D. melanogaster*, such as the detection of sex pheromone, salt, and water (Liu et al. 2003; Cameron et al. 2010; Liu et al. 2012). TRPs are a highly conserved cation channel family playing diverse sensory and non-sensory roles. A few members of TRPs, such as TRPA1 and TRPM, are involved in sensing noxious chemicals which trigger escaping behavior (Kang et al. 2010; Himmel et al. 2019). Finally, a family of small secreted proteins, called odorant binding proteins (OBPs), also play important roles in the sensory transduction (Sun et al. 2018). OBPs are present in sensillum lymph surrounding sensory neurons, and are generally thought to mediate the transport of hydrophobic ligands to chemoreceptor proteins in an aqueous environment (Sun et al. 2018; Schmidt and Benton 2020). Functional characterization of these six chemosensory gene families have been the central focus of insect chemosensory research in the past two decades.

Facilitated by the recent innovation of genome sequencing technologies, evolutionary dynamics of the chemosensory gene families have been investigated in several insect lineages, which in turn gives insight into the genetic architecture of behavioral evolution (Robertson 2019). Lineage-specific gene duplications or gene losses are the most prevalent mode of evolution, especially for the *Ors* and *Grs (Robertson 2019).* In some cases, gene turnover events of those two gene families were associated with the chemosensory adaptation in ecological specialists (McBride 2007; McBride and Arguello 2007; Obiero et al. 2014; Goldman-Huertas et al. 2015; Matsunaga et al. 2022). Evolution of ligand specificity in orthologous chemoreceptors has also been illuminated in several species, often coupled with the acceleration of evolutionary rates (e.g., dN/dS) or the identification of amino acid residues underlying ligand tunings (Prieto-Godino et al. 2017; Yang et al. 2018; Auer et al. 2020; Brand et al. 2020; Li et al. 2023).

In contrast to the two modes of molecular evolution mentioned above, transcriptional evolution in chemosensory gene families – and how the gene expression change leads to the novel behavioral output – has been gaining less attention. This lack of knowledge is because of technical difficulties in current differential expression (DE) analysis frameworks assuming similar gene expression profiles between samples compared (Zhou et al. 2019; Price et al. 2022). As a result, applying interspecies comparisons to this framework yields many false positives due to the differences in gene lengths and gene expression depths (Zhou et al. 2019; Price et al. 2022). In fact, the best example in this vein is derived from a within-species comparison of the yellow fever mosquito, *Aedes aegypti*, wherein an expression difference in a single *Or* gene encoding a receptor for a human-enriched odorant explains the population differences in host preference for blood-feeding (McBride et al. 2014). Despite the statistical uncertainty, there have been several attempts to apply DE analyses to cross-species comparisons in the genus *Drosophila*, with emphasis on ecological specialists like *D.*

sechellia (Shiao et al. 2015) or *D. suzukii* (Dweck et al. 2021; Wang et al. 2022). Notably, Dweck et al. (2021) reduced false positives from interspecific DE analyses by taking a consensus of four different pipelines, which could be easily applicable to other pairwise comparisons. A recent study further expanded the analysis to a phylogenetic scale containing 5-6 species by treating TPM (transcripts per million, (Wagner et al. 2012)) values of orthologous genes as continuous trait values (Bontonou et al. 2024). All of these past studies successfully illuminated the candidate genes underlying chemosensory evolution in drosophilids, highlighting gene expression divergence as a promising mechanism of behavioral evolution.

Previous researches on *S. flava* convincingly showed that chemosensory gene families indeed played major roles in behavioral evolution via gene losses (Goldman-Huertas et al. 2015), gene gains (Matsunaga et al. 2022), and acceleration of protein evolution (Peláez et al. 2023). However, contributions of gene expression in these families have not been investigated yet. In this chapter, we hypothesized that the chemosensory evolution in the herbivorous *S. flava* lineage has been partly driven by the transcriptional evolution of the chemosensory gene families, and under this assumption, we comprehensively searched for the candidate genes that might have contributed to the ecological adaptation of *S. flava* via gene expression changes. Our basic strategy was to conduct bulk RNA-sequencing in *S. flava* and its closest microbe-feeding relative with a reference genome, *S. pallida*, to perform interspecific comparisons of transcriptome profiles. To thoroughly capture chemosensory gene expressions, we sequenced samples from both sexes and four major chemosensory organs of the adult fly: Antenna, maxillary palp, labellum, and tarsus (Fig. 1A). Antenna and maxillary palp are primarily olfactory organs, while labellum and tarusi are considered gustatory organs. Using these RNA-seq datasets, we conducted three different types of interspecies comparisons of gene expression: 1) RPM (reads per million)-based analyses of gains/losses of chemosensory gene expression, 2) RPM-based comparisons of gene expression parameters, Euclidean distance and tau (τ), a parameter of spatial expression specificity (Yanai et al. 2005), between orthologous chemosensory gene pairs, 3) Count-based DE analyses of targeting the entire gene set. These complementary approaches together highlighted key candidate genes, or important characteristics of the genes, that could have facilitated the chemosensory evolution in this charismatic fly lineage.

Materials and methods

Fly strains. The *S. flava* lab colony was originally established from >150 wild larvae collected from Dover, NH (USA) in 2008. The colony was maintained in cages containing *A. thaliana* Col-0 strain as larval and adult food source at 20-21°C and 60% humidity in a

walk-in environmental room on a 12L:12D cycle. The *S. pallida* lab colony was an isofemale line established from a wild isofemale fly line collected at the University of California, Berkeley campus (CA, USA) in 2017. The colony was reared on organic spinach (purchased at Trader Joe's, CA, USA) placed on top of regular cornmeal media purchased from UC-Berkeley Fly Food Facility. Other rearing conditions were identical to those of the *S. flava* colony.

Sample collection. Newly emerged flies of each species were collected from the cages/vials soon after the eclosion and kept in humidified vials containing 10% honey water, aiming to minimize their larval food differences. 3-10 day-old flies were anesthetized on CO_2 and each chemosensory organ was hand-dissected using razors and forceps as described below: For antennae, second and third antennal segments (including arista) were carefully removed by fine forceps. Maxillary palps were also collected by fine forceps, and labella were collected after the maxillary palp removal to avoid cross-tissue contamination. Tarsal segments were removed by razors from all six legs and pooled together. Dissected tissues were directly collected in LB+TGA lysis buffer from Reliaprep RNA Tissue Miniprep System (Promega, USA) and crushed using Biomasher Standard homogenizer (Takara Bio Inc., USA) in a dry ice ethanol bath. Sample lysates were stored in -80°C freezer until subsequent RNA extraction.

RNA extraction and bulk RNA-sequencing. Approximate numbers of individuals pooled for a single replicate were 200-300 for antennae, 100-120 for maxillary palps and labella, and 70-75 for tarsi. Total RNA were extracted from the lysates using ReliaPrep RNA Tissue Miniprep System (Promega, USA) according to the manufacturer's protocol, and quantified using Qubit RNA High Sensitivity kit (Thermo Fisher Scientific). Extracted RNA samples from antennae were subsequently processed into cDNA libraries using KAPA RNA HyperPrep Kit (Roche Sequencing) at Functional Genomics Laboratory at UC Berkeley. Labellar and tarsal samples were prepared by HS according to the manufacturer's protocol. Maxillary palp RNA samples suffered from low RNA yields, therefore we employed low-input library prep methods (SMART-seq) which were performed at Functional Genomics Laboratory at UC Berkeley. All samples were sequenced on NovaSeq 6000 platforms (Illumina, USA) for 150bp paired-end reads.

Gene annotation and read mapping. We used previously reported reference genome assemblies for *S. flava* (Peláez et al. 2023) and *S. pallida* (Kim et al. 2021) for subsequent bioinformatic analyses. Although gene annotations for the two reference genomes were produced previously, we noticed that there are several occasions in which CDS structures and gene boundaries were incorrectly identified in these initial annotations. Therefore, we began our bioinformatic pipeline by updating CDS

annotations by the novel gene prediction using the latest RNA-sequencing data and the manual inspection (Fig. 1C). We first filtered raw RNA-seq reads using Fastp v0.21.0 (Chen et al. 2018), and mapped the clean reads on respective reference genomes using STAR v2.7.1a (Dobin et al. 2013) to generate multiple alignment (BAM) files. Next, we ran BRAKER v2 (Hoff et al. 2016; Brůna et al. 2021) to generate new gene annotation files (.gtf) using two different inputs in parallel; whole protein sequences from *D. melanogaster* genome v6.45, and our newly generated BAM files from respective *Scaptomyza* species. Protein-derived and BAM-derived annotations achieved comparable gene coverages (i.e., percentages of complete BUSCO genes were 97.2% in protein-derived and 97.0% in BAM-derived annotation in *S. flava*), but some genes were still missed out in one of the two annotation types. To complement those lacks, we used GffCompare v0.11.2 (Pertea and Pertea 2020) to merge the two annotation types to create a single annotation file that has the best gene coverage. At this point, however, we recognized that some genes were still incorrectly annotated, particularly at translation start sites and exon-intron boundaries. Since our main interest was gene expression of chemosensory genes, we manually inspected the annotations of six major chemosensory gene families (i.e., *Or, Gr, Ir, Obp, Ppk,* and *Trp*) to correct their gene coordinates if necessary by referring to previously used gene models (Peláez et al. 2023) and BAM files (Fig. 1C). We used these manually corrected annotations for subsequent conversion of read alignments into count data using HTseq v0.9.1 (Anders et al. 2015).

Orthology inference. We performed reciprocal BLASTN searches using the whole CDS sets of *S. flava* and *S. pallida* to identify 1:1 orthologous relationships between species using manual Python scripts, resulting in 10,439 orthologous gene pairs. Genes experienced lineage-specific duplications were retained in the orthologue set as long as its counterpart was correctly identified and formed a best-hit pair that did not overlap with any other pairs (i.e., paralogues tied to the same gene with lower e-values/bitscores were omitted from the list). Using this dataset, we conducted principal component analysis (PCA) to visualize similarities among samples using prcomp package in R v4.2.1. Note that we used this 1:1 orthologue datasets for PCA and DEG analyses, while we included duplicated genes in the gene expression heatmaps and the computation of gene expression parameters.

Expression heatmaps, analysis of expression gains/losses. Count data were converted to RPM by normalizing the counts with the number of total reads mapped on all annotated genes for each sample, and transformed to the Log_2 scale. We associated gene expression data to each chemosensory gene using a manual Python script, and visualized the gene expression levels using pheatmap package in R v4.2.1. Genes

experienced recent lineage-specific duplications (i.e., either in the herbivorous *Scaptomyza* or the *S. pallida* lineage) were identified in Pelaez et al. (2023), and such genes were highlighted on the heatmaps. Based on these datasets, we also counted the number of genes with sufficiently high expression (i.e., Mean(Log2[RPM+1]) > 1) and the proportions out of the entire gene family in each tissue.

Expression divergence metrics. We used two different pairwise metrics of gene expression divergence, Euclidean distance (Liao and Zhang 2006) and tissue expression specificity, tau (τ) (Yanai et al. 2005). The former provides a simple geometric distance between two profiles, while the latter takes into account an expression heterogeneity among tissues for a given gene. Here, we used log-transformed gene expression values (i.e., $\log_2$[RPM+1]), and treated male and female organs as different "tissue" to take into account sex-specific gene expression. Calculation of Euclidean distance was based on average expression levels of three replicates in each tissue, while in τ each replicate was considered separately to incorporate within-sample heterogeneity into the metrics. Variations in the two parameters among gene families were analyzed by Kruskal-Wallis test, followed by post-hoc tests using Dunn's test with Benjamini-Hochberg correction. Differences between single-copy and duplicated genes were tested by Wilcoxon rank sum test. Both tests were performed in R v4.2.1.

Differential gene expression analyses. Cross-species differential expression (DE) analysis is statistically underdeveloped and intrinsically suffers from a high rate of false positives. To address this issue, we employed three independent DE analysis pipelines and took the consensus of all three to narrow down more plausibly differentially expressed genes (DEGs) (Fig. 1B), referring to a similar strategy used in (Dweck et al. 2021).

To search for differentially expressed genes from 10,439 orthologous gene pairs, we leveraged three differential expression analysis software programs: DESeq2 (Love et al. 2014), edgeR (Robinson et al. 2010), and SCBN + NOIseq (Tarazona et al. 2015; Zhou et al. 2019) (Fig. 1B). SCBN normalizes raw count data by taking the interspecific differences in gene lengths and gene expression depths designed for cross-species expression analysis (Zhou et al. 2019). Since SCBN itself does not perform DE analyses, we input the normalized count data to NOIseq for statistical analysis. To minimize the chance of getting false positives, genes were filtered with a strict cutoff, $|\text{Log}_2\text{FoldChange}| > 2$ and adjusted p-value < 0.001, in order to be designated as DEGs for each DE analysis pipeline. Only genes that satisfied the DE criteria in all three pipelines were treated as "true" DEGs for subsequent analysis (Fig. 1B). We then investigated the overlap between DEGs from a female-comparison and a male-comparison in a single chemosensory organ to categorize the genes as

"female-specific", "male-specific", and "universal" DEGs (Fig. 1B). We also investigated and visualized the overlaps of DEGs between the four organs using the Intervene web platform (Khan and Mathelier 2017). We also performed GO term enrichment analysis for each DEG category by the topGO package using the "elim" algorithm (Alexa 2006).

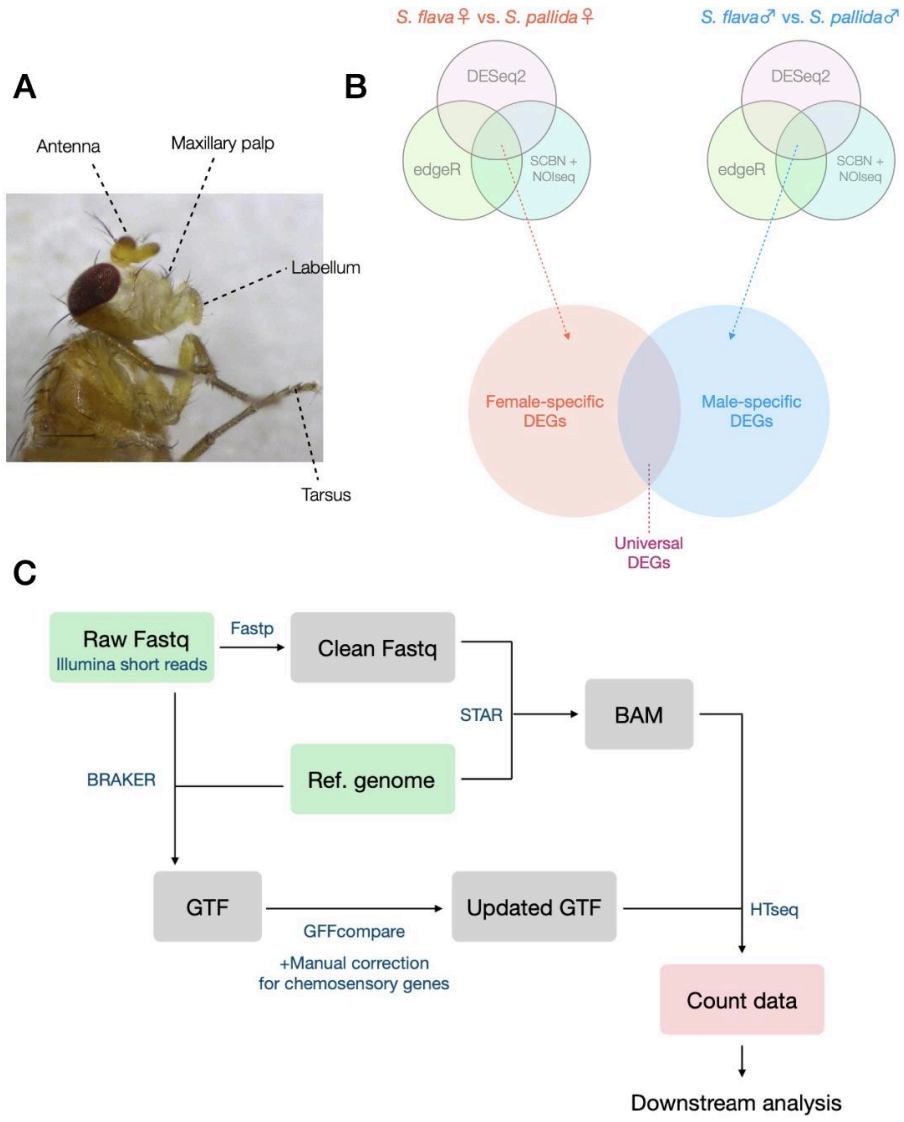


Figure 1. Analytical frameworks of Chapter 2. (A) A female *S. flava* highlighting the four chemosensory organs targeted in this study (i.e., antenna, maxillary palp, labellum, and tarsus). Tarsal samples were derived from all six legs of adult flies. (B) Identification of differentially expressed genes (DEGs) in this study. We first conducted interspecies/same-sex DEG analysis using three pipelines (i.e., DESeq2, edgeR, and SCBN/NOIseq), took the consensus, and treated those genes as "true" DEGs in a given comparison. Subsequently, we compared the repertoire of DEGs between sexes to identify "sex-specific" and "universal" DEGs in an organ. (C) Bioinformatics pipeline of this study (see Methods section for more information).

Results

Broad gene expression patterns in PCA plots. We sequenced 48 samples in total (i.e., 16 categories (2 species/2 sexes/4 organs) with 3 biological replicates for each). Principal component analysis (PCA) for 10,439 genes with 1:1 orthologous gene pairs revealed that the samples were roughly clustered by organs of origin, especially for labellum and tarsus (Fig. 2A, 2B). The first principal component (PC1) generally separated the clusters of antenna, labellium/tarsus, and maxillary palps, whereas the second principal component (PC2) separates *S. flava* or *S. pallida*-derived samples (Fig. 2A). The third principal component (PC3) roughly explains the difference between olfactory (antenna and maxillary palp) and gustatory organs (labellum and tarsus) (Fig. 2B). Antennal samples were rather scattered around the plot in which male *S. pallida* samples appear as outliers in the lower right of Fig. 2A. We hypothesize that this pattern could be attributed to the differences in sample collections and library prep procedures and does not necessarily reflect real biological differences. Therefore, our results from gene expression analysis in antennal samples should be interpreted somewhat cautiously with respect to *S. pallida*.

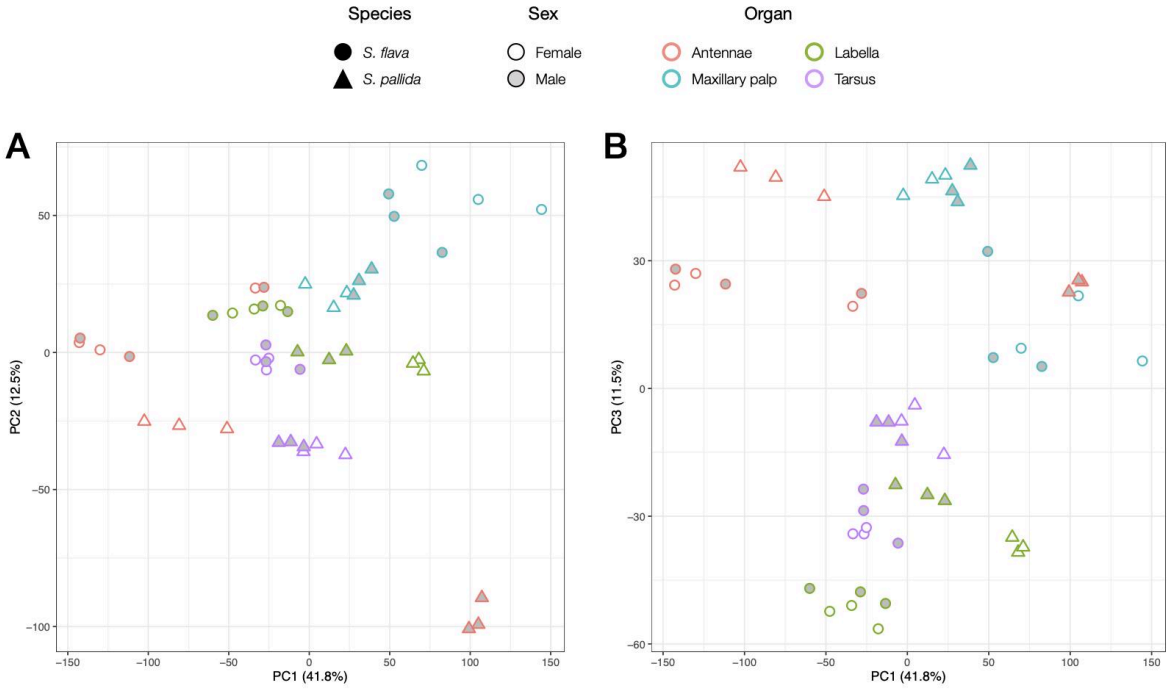
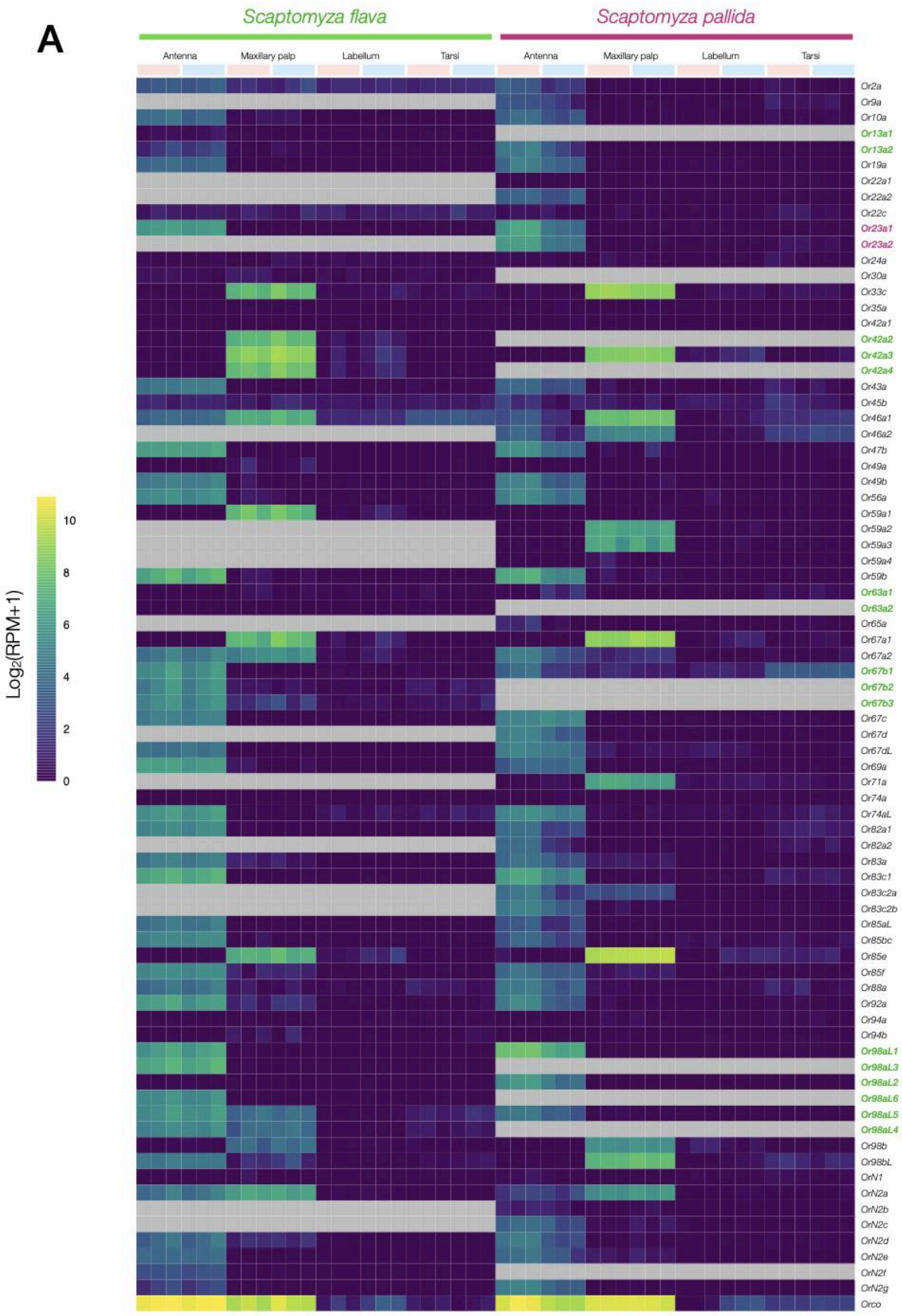
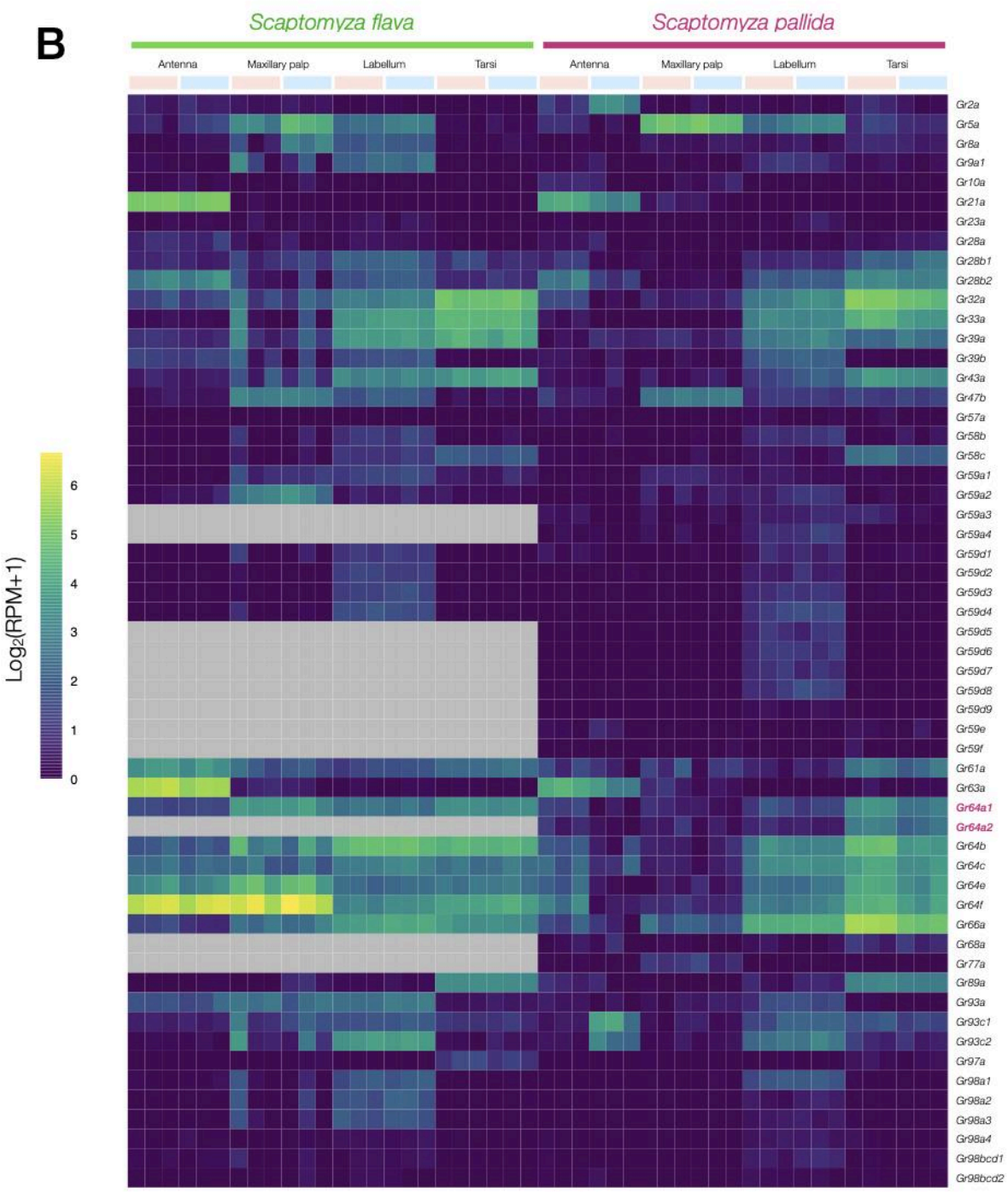
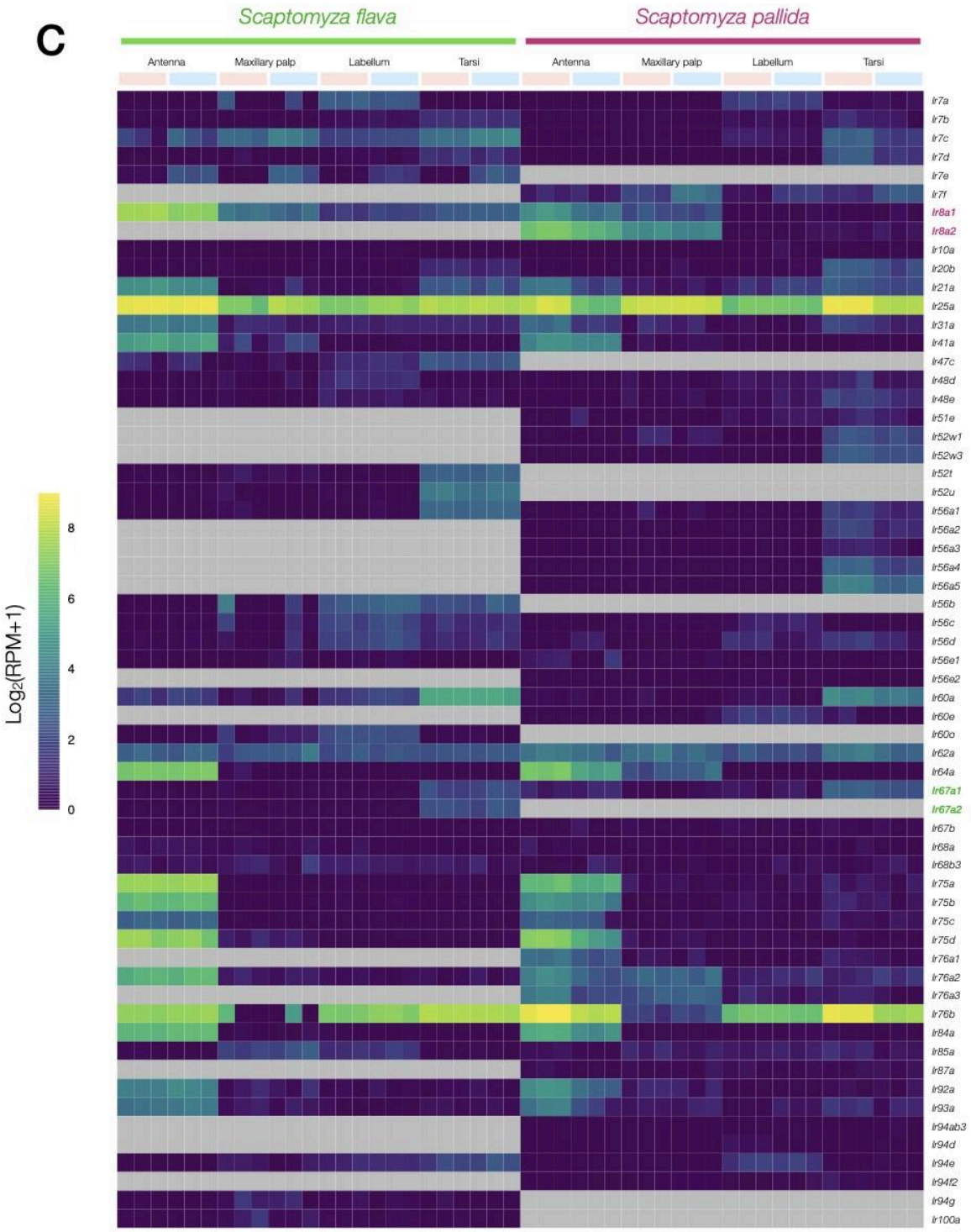


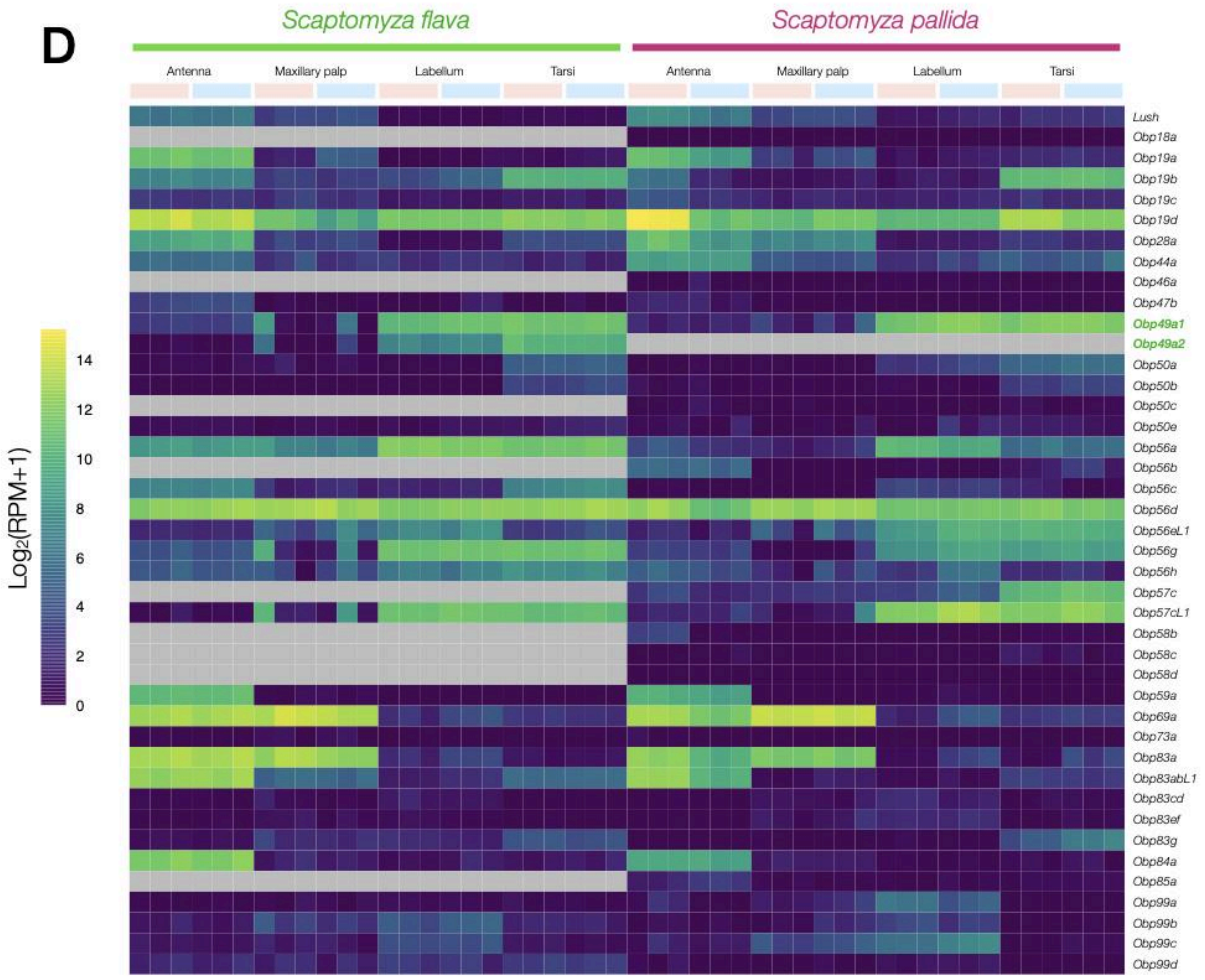
Figure 2. PCA plots for chemosensory transcriptomes of *S. flava* and *S. pallida*. (A) A plot for PC1 and PC2 axes. (B) A plot for PC1 and PC3 axes. Values in brackets show percentages of variance explained by each principal component.

Chemosensory gene expression atlas. We next visualized the expression levels of all annotated chemosensory genes in heatmaps to get a better view of expression divergence between species. Despite several clade-specific gene gains or losses across the phylogeny (Peláez et al. 2023), general expression profiles of chemosensory gene families were largely conserved, because the numbers of sufficiently expressed genes (i.e., log2(RPM+1) > 1) in each organ, or the proportions of those genes, were fairly comparable between species (Table 1A, 1B). A few exceptions to this pattern were *Grs* expressed in maxillary palp and labellum, where a higher proportion of genes were expressed in *S. flava* (Table 1B). However, some maxillary palp samples exhibited strange heterogeneity in gene expression potentially indicating contamination from labellar samples (e.g., Fig. 2B). Thus, some results from our maxillary palp analysis should be taken cautiously.


A
Scaptomyza flava
Scaptomyza pallida
Antenna
Maxillary palp
Labellum
Tarsi
Antenna
Maxillary palp
Labellum
Tarsi
Log2(RPM+1)
10
8
6
4
2
0
Or2a
Or9a
Or10a
Or13a1
Or13a2
Or19a
Or22a1
Or22a2
Or22c
Or23a1
Or23a2
Or24a
Or30a
Or33c
Or35a
Or42a1
Or42a2
Or42a3
Or42a4
Or43a
Or45b
Or46a1
Or46a2
Or47b
Or49a
Or49b
Or56a
Or59a1
Or59a2
Or59a3
Or59a4
Or59b
Or63a1
Or63a2
Or65a
Or67a1
Or67a2
Or67b1
Or67b2
Or67b3
Or67c
Or67d
Or67dL
Or69a
Or71a
Or74a
Or74aL
Or82a1
Or82a2
Or83a
Or83c1
Or83c2a
Or83c2b
Or85aL
Or85bc
Or85e
Or85f
Or88a
Or92a
Or94a
Or94b
Or98aL1
Or98aL3
Or98aL2
Or98aL6
Or98aL5
Or98aL4
Or98b
Or98bL
OrN1
OrN2a
OrN2b
OrN2c
OrN2d
OrN2e
OrN2f
OrN2g
Orco


B
Scaptomyza flava
Scaptomyza pallida
Antenna
Maxillary palp
Labellum
Tarsi
Antenna
Maxillary palp
Labellum
Tarsi
Log2(RPM+1)
0
1
2
3
4
5
6
Gr2a
Gr5a
Gr8a
Gr9a1
Gr10a
Gr21a
Gr23a
Gr28a
Gr28b1
Gr28b2
Gr32a
Gr33a
Gr39a
Gr39b
Gr43a
Gr47b
Gr57a
Gr58b
Gr58c
Gr59a1
Gr59a2
Gr59a3
Gr59a4
Gr59d1
Gr59d2
Gr59d3
Gr59d4
Gr59d5
Gr59d6
Gr59d7
Gr59d8
Gr59d9
Gr59e
Gr59f
Gr61a
Gr63a
Gr64a1
Gr64a2
Gr64b
Gr64c
Gr64e
Gr64f
Gr66a
Gr68a
Gr77a
Gr89a
Gr93a
Gr93c1
Gr93c2
Gr97a
Gr98a1
Gr98a2
Gr98a3
Gr98a4
Gr98bcd1
Gr98bcd2


C
Scaptomyza flava
Scaptomyza pallida
Antenna
Maxillary palp
Labellum
Tarsi
Antenna
Maxillary palp
Labellum
Tarsi
Ir7a
Ir7b
Ir7c
Ir7d
Ir7e
Ir7f
Ir8a1
Ir8a2
Ir10a
Ir20b
Ir21a
Ir25a
Ir31a
Ir41a
Ir47c
Ir48d
Ir48e
Ir51e
Ir52w1
Ir52w3
Ir52t
Ir52u
Ir56a1
Ir56a2
Ir56a3
Ir56a4
Ir56a5
Ir56b
Ir56c
Ir56d
Ir56e1
Ir56e2
Ir60a
Ir60e
Ir60o
Ir62a
Ir64a
Ir67a1
Ir67a2
Ir67b
Ir68a
Ir68b3
Ir75a
Ir75b
Ir75c
Ir75d
Ir76a1
Ir76a2
Ir76a3
Ir76b
Ir84a
Ir85a
Ir87a
Ir92a
Ir93a
Ir94ab3
Ir94d
Ir94e
Ir94f2
Ir94g
Ir100a
Log2(RPM+1)
8
6
4
2
0


D
Scaptomyza flava
Scaptomyza pallida
Antenna
Maxillary palp
Labellum
Tarsi
Antenna
Maxillary palp
Labellum
Tarsi
Log2(RPM+1)
14
12
10
8
6
4
2
0
Lush
Obp18a
Obp19a
Obp19b
Obp19c
Obp19d
Obp28a
Obp44a
Obp46a
Obp47b
Obp49a1
Obp49a2
Obp50a
Obp50b
Obp50c
Obp50e
Obp56a
Obp56b
Obp56c
Obp56d
Obp56eL1
Obp56g
Obp56h
Obp57c
Obp57cL1
Obp58b
Obp58c
Obp58d
Obp59a
Obp69a
Obp73a
Obp83a
Obp83abL1
Obp83cd
Obp83ef
Obp83g
Obp84a
Obp85a
Obp99a
Obp99b
Obp99c
Obp99d

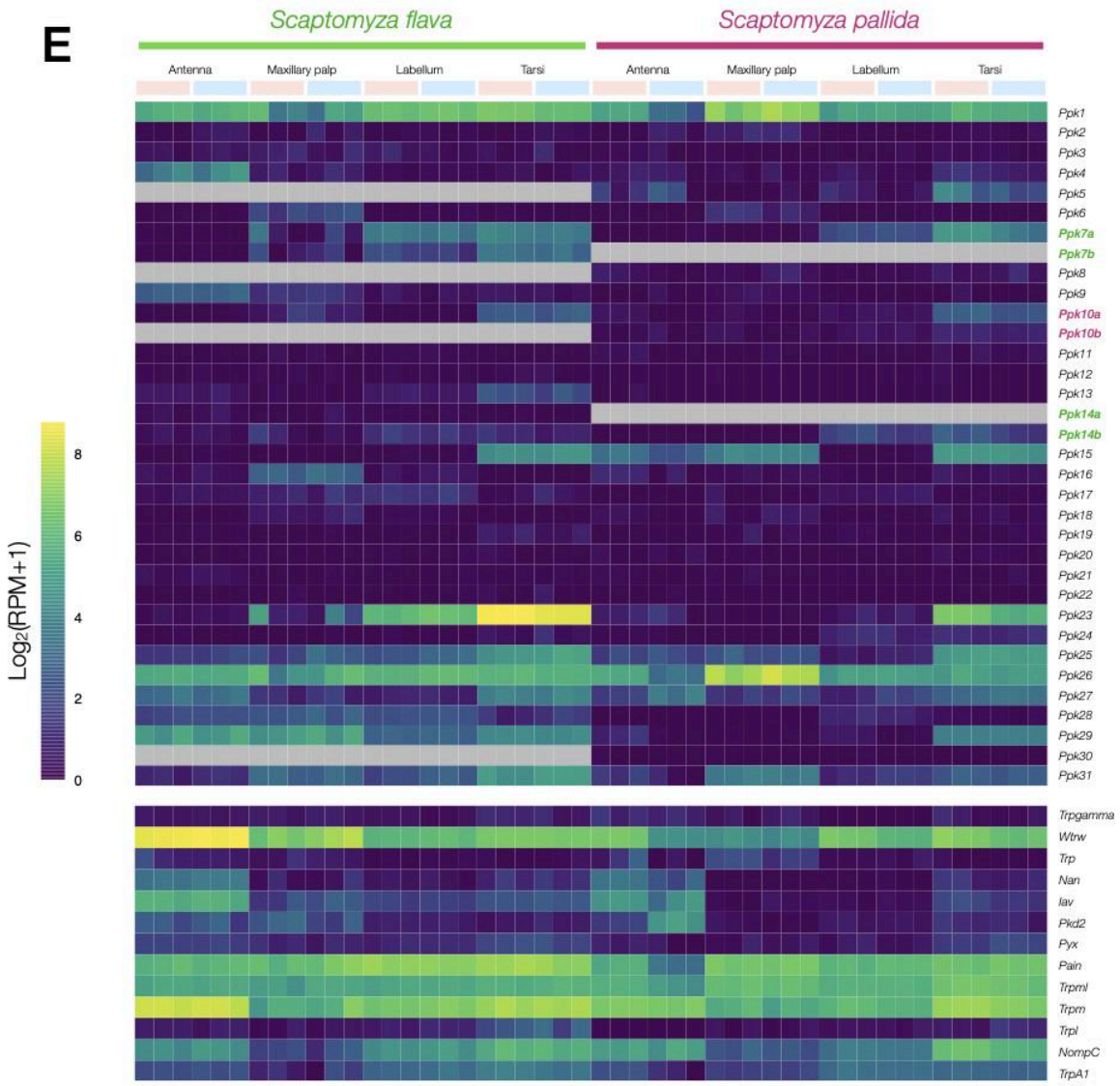


Figure 3. Gene expression atlases for chemosensory gene families. (A-E) Log-transformed RPM values are plotted as heatmaps for (A) Olfactory receptor (*Or*), (B) Gustatory receptor (*Gr*), (C) Ionotropic receptor (*Ir*), (D) Odorant-binding protein (*Obp*), (E) Pickpocket (*Ppk*) and Transient Receptor Potential (*Trp*) gene families. Three biological replicates were sequenced for each sex/organ/species. Pale-colored bars at the top indicate sexes of the samples (i.e., Pale red: female, Pale blue: male). Genes experienced a recent lineage-specific duplication event were highlighted in colored gene names (Green: *S. flava*-specific or herbivore-specific gene duplication, Purple: *S. pallida*-specific duplication. Information on gene gains and losses were taken from (Peláez et al. 2023))).

(A)

Family	*S. flava*									*S. pallida*								
	total	Antenne		Max. palp		Labellum		Tarsus		total	Antenna		Max. palp		Labellum		Tarsus	
		F	M	F	M	F	M	F	M		F	M	F	M	F	M	F	M
Or	62	39	40	19	21	3	7	2	4	67	45	43	16	15	0	4	8	5
Gr	44	14	13	21	19	31	31	16	14	56	13	7	5	4	21	25	19	18
Ir	44	18	19	8	8	13	16	19	20	52	19	19	9	9	8	9	22	17
Obp	33	22	21	24	23	20	20	21	23	41	26	23	16	22	19	25	22	22
Ppk	30	8	9	12	13	11	11	14	13	32	9	6	7	6	7	6	14	13
Trp	13	11	10	8	8	7	9	11	10	13	11	8	7	7	6	6	10	9

(B)

Family	Difference in %expressed genes							
	Antenna		Maxillary palp		Labellum		Tarsus	
	F	M	F	M	F	M	F	M
Or	0.043	0.003	0.068	0.115	0.048	0.053	0.087	0.01
Gr	0.086	0.17	0.388	0.36	0.33	0.258	0.024	0.003
Ir	0.044	0.066	0.009	0.009	0.142	0.191	0.009	0.128
Obp	0.033	0.075	0.337	0.16	0.143	0.004	0.1	0.16
Ppk	0.015	0.113	0.181	0.246	0.148	0.179	0.029	0.027
Trp	0	0.154	0.077	0.077	0.077	0.231	0.077	0.077

Table 1. Numbers and proportions of expressed chemosensory genes. (A) Numbers of sufficiently expressed genes (i.e., log2(RPM+1)>1) across the four chemosensory organs in *S. flava* and *S. pallida*. (B) Interspecific differences in the proportion of expressed genes at each category. Tissue/gene family with top 5 differences are highlighted in blue.

Expression divergence in orthologous pairs. We further addressed whether and to what extent the pattern of gene expression divergence in an orthologous pair varies between gene families or history of recent gene duplications. Here, we used two quantitative metrics, Euclidean distances (Liao and Zhang 2006) and differences in tissue specificity index, tau (τ) (Yanai et al. 2005), which we called Δτ, to quantitatively analyze expression differences between species. The former provides a simple geometric distance between two profiles, while the latter takes into account an expression heterogeneity among tissues for a given gene.

Comparisons of Euclidean distance revealed that *Obps* are the only gene family that stood out of all other families (Fig. 4A), while interspecific differences in tissue specificity (Δτ) were highest among *Ppks* (Fig. 4C). These inconsistencies between metrics likely arose from the fact that Euclidean distance was not necessarily normalized by absolute expression levels, which fluctuated in genes with higher overall expression, like *Obps*. Although lacking statistical significance, the *Ppk* and *Obp* gene families ranked the second highest group for Euclidean distance and Δτ, respectively (Fig. 4A, 4C). On the other hand, we did not find significant differences in these metrics between single-copy and duplicated gene pairs (Fig. 4B, 4D). Overall, our analysis highlighted the variability of evolutionary dynamics of gene expression across gene families.

Some of the top-ranked genes in Δτ included functionally annotated genes in *D. melanogaster*, such as *Ppk29* (sex pheromone detection; (Liu et al. 2012; Thistle et al. 2012)), *Ppk28* (water detection; (Cameron et al. 2010)), and *Ir7c* (high salt detection; (McDowell et al. 2022)). Intriguingly, all of these genes have lost their tissue specificity in *S. flava* (i.e., expressed in a wider variety of organs) compared to the microbe-feeding *S. pallida* (Fig. 3C, 3E).

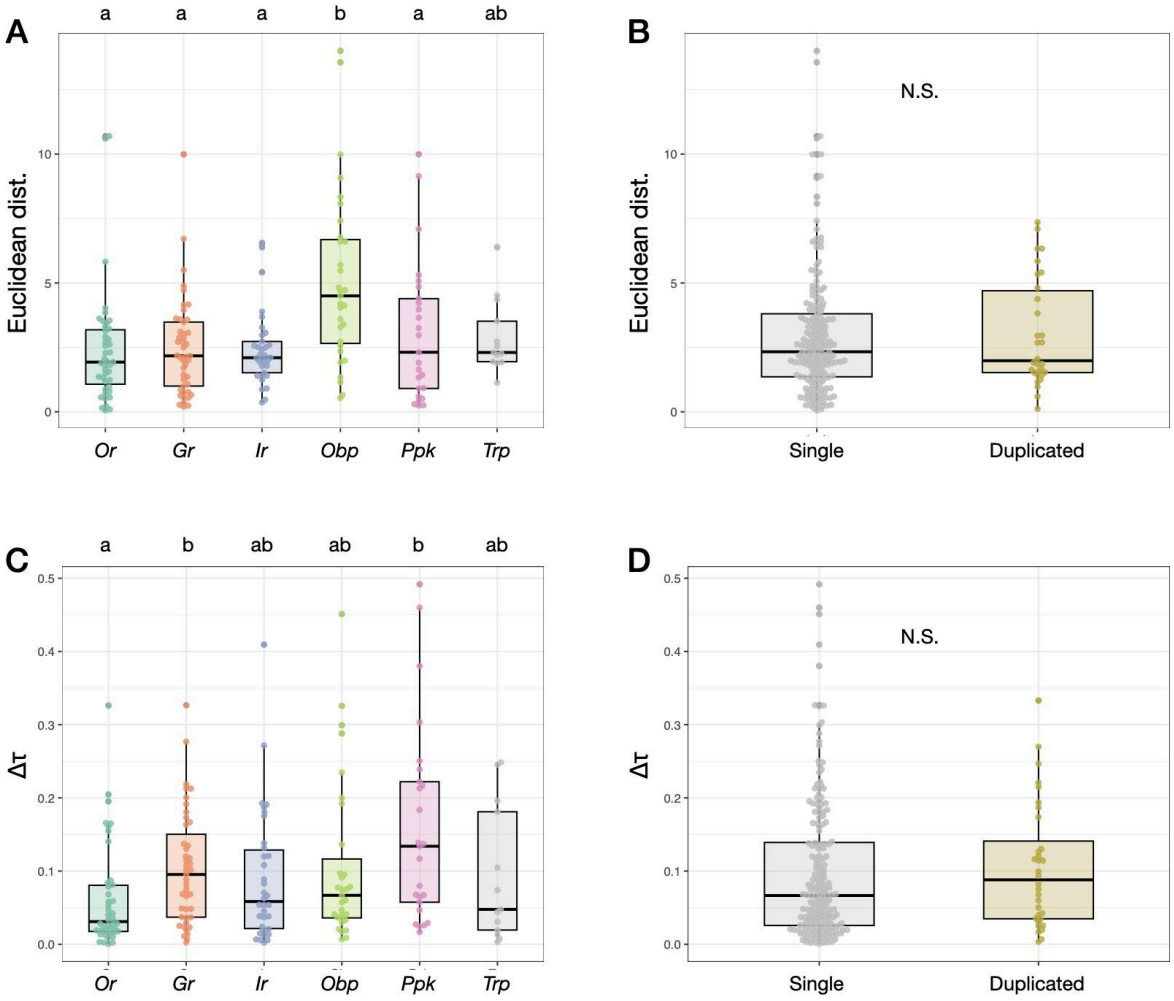


Figure 4. Gene expression divergence in orthologous pairs between *S. flava* and *S. pallida*. (A, B) Euclidean distance plotted for (A) gene families and (B) recent history of gene duplication. (C, D) Pairwise difference in tissue specificity of gene expression, Δτ, plotted for (C) gene families or (D) single-copy and duplicated pairs. Analyses of gene families were only performed for single-copy pairs. Variations among gene families were analyzed by Kruskal-Wallis test, followed by a post-hoc test using Dunn's test with Benjamini-Hochberg correction. Letters show the significant difference between families. Differences between single-copy and duplicated genes were tested by Wilcoxon rank sum test.

Interspecific differential expression analysis. A series of DE analysis pipelines revealed comparable numbers of cross-species DEGs across organs, ranging between 505 in male maxillary palp and 1,454 in male antenna (Fig. 5). We performed GO enrichment analysis in each group and, as expected, we found significant enrichment of GO terms related to chemosensation in several comparisons (Table 2). Indeed, many chemosensory genes were annotated as DEGs in all comparisons. Some genes appeared as DEGs in more than four sex-specific comparisons, including functionally annotated genes like *Or67b1, Ir7c,* and *Ppk23* (Fig. 5). Among the six gene families, the proportion of genes differentially expressed at least in one comparison was highest in the *Ppk* (53.6%), followed by the *Obp* (45.2%) families (Table 3).

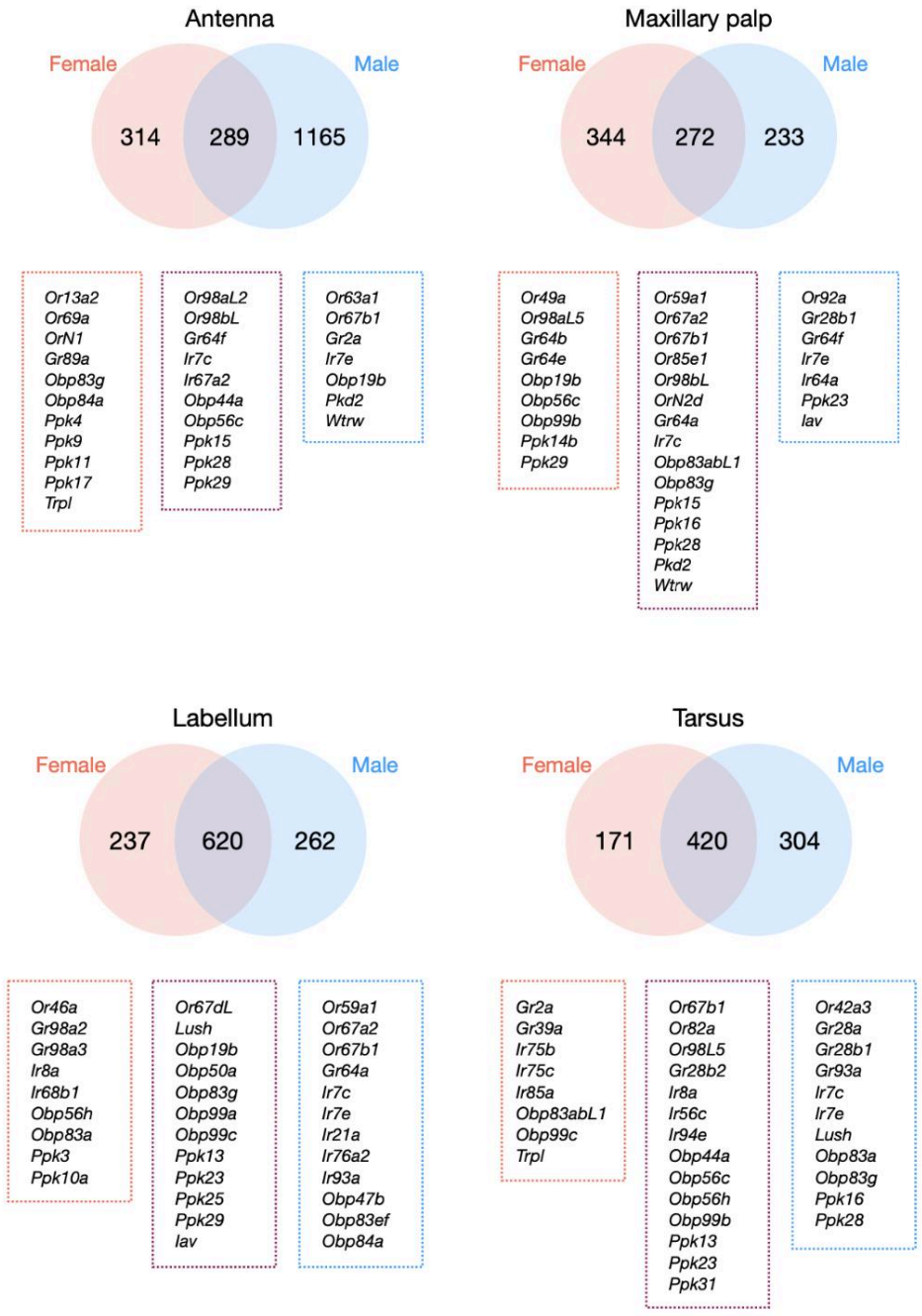


Figure 5. Differentially expressed genes between *S. flava* and *S. pallida* in four chemosensory organs. Venn diagrams show the numbers of female-specific, male-specific, and sexually universal DEGs in each organ between *S. flava* and *S. pallida*. Chemosensory genes included in each category were highlighted in the boxes below.

Tissue	Sex	ID	GO Term	#Annotated	#Significant	#Expected	p-value
antennae	universal	GO:0019233	sensory perception of pain	112	7	2.39	0.0099
antennae	female	GO:0007606	sensory perception of chemical stimulus	252	14	5.59	0.0014
maxillary_palp	universal	GO:0009593	detection of chemical stimulus	181	11	3.7	0.0012
maxillary_palp	universal	GO:0009415	response to water	15	3	0.31	0.0032
maxillary_palp	universal	GO:0007608	sensory perception of smell	164	9	3.35	0.00646
maxillary_palp	female	GO:0007605	sensory perception of sound	147	10	3.81	0.0048
maxillary_palp	female	GO:0050953	sensory perception of light stimulus	90	7	2.33	0.00863
maxillary_palp	female	GO:0030534	adult behavior	264	14	6.84	0.00867
maxillary_palp	male	GO:0050909	sensory perception of taste	73	5	1.25	0.00806
maxillary_palp	male	GO:0007619	courtship behavior	107	6	1.83	0.0099
labella	male	GO:0009593	detection of chemical stimulus	181	11	3.31	0.00047
labella	male	GO:0009636	response to toxic substance	461	18	8.43	0.00187
tarsi	female	GO:0007600	sensory perception	582	15	7.34	0.0063
tarsi	male	GO:0040040	thermosensory behavior	20	3	0.44	0.00905

Table 2. Significant GO term enrichments in interspecific DEGs. GO terms associated with sensation or behavioral traits were listed. Terms related to chemoreception were highlighted in blue.

Family	#Total orthologous pairs	#Pairs with DE in at least one comparison	% DEGs
Or	51	18	35.3
Gr	44	13	29.5
Ir	35	13	37.1
Obp	31	14	45.2
Ppk	28	15	53.6
Trp	13	4	30.8

Table 3. Numbers and percentages of DEGs in the six chemosensory gene families. GO terms associated with sensation or behavioral traits were listed. Terms related to chemoreception were highlighted in blue.

Patterns of universal DEGs across chemosensory organs. Interspecific DEGs shared between sexes, which we call "universal DEGs", are more likely associated with the differences between mustard-feeders and microbe-feeders, and are likely decoupled from sex-specific behavior or physiology. We found that several chemosensory genes categorized as universal DEGs in two different organs, including ones previously named in other analyses (e.g., *Or67b1*, *Ir7c*, *Ppk23*) (Fig. 6A). No chemosensory genes were differentially expressed in more than three organs (Fig. 6A), although some genes still showed up as sex-specific DEGs.

UpSet plot showed that the majority of universal DEGs were organ-specific (Fig. 6B). Still, we found 60 genes that were differentially expressed between species in all organs (Fig. 6B). These highly universal DEGs included neither chemosensory genes nor chemosensation-related GO enrichment, but there are a few genes that might be associated with the mustard feeding ecology, such as *GstE2*, a member of the Glutathione-*S*-transferase (*Gst*) gene family, which was drastically downregulated in *S. flava*.

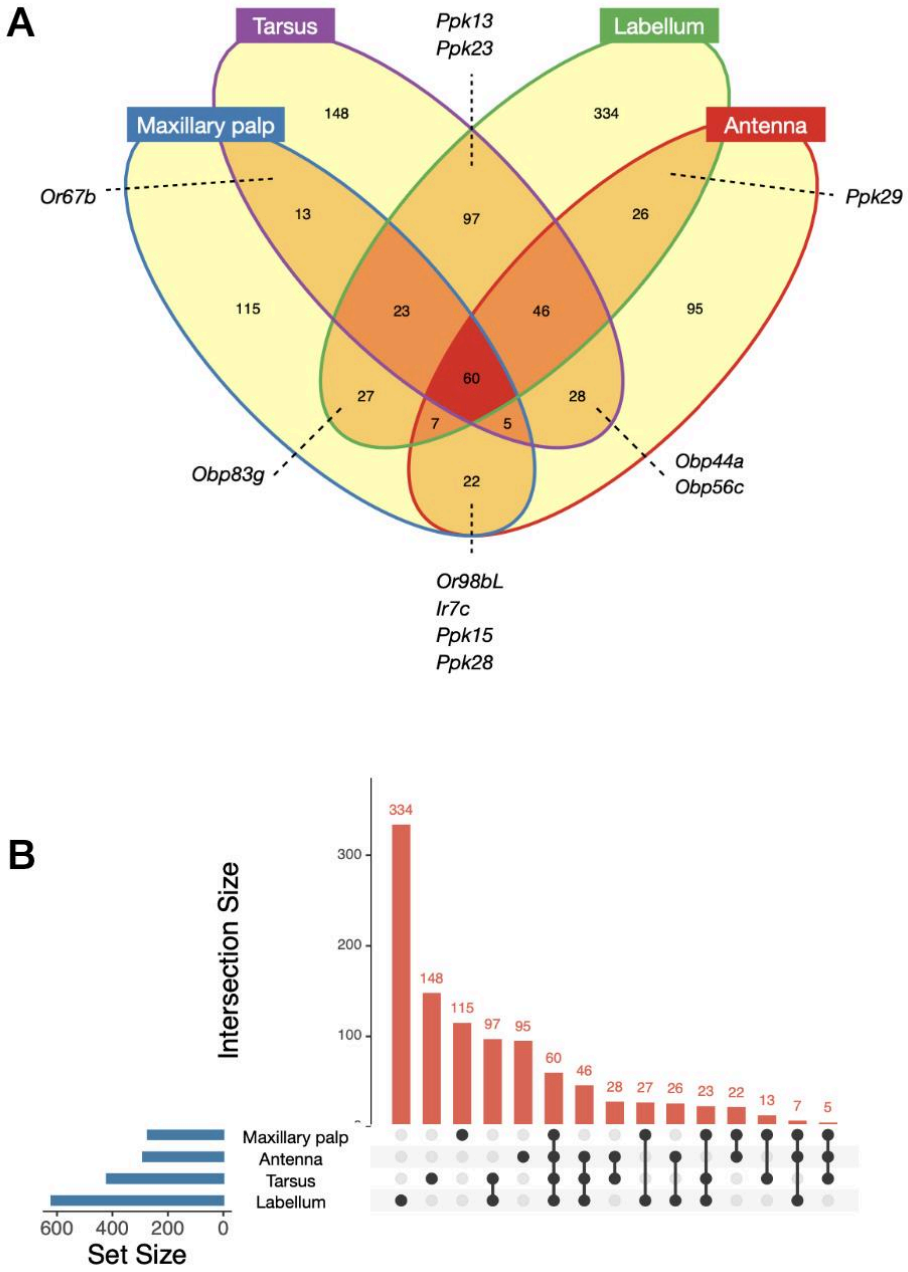


Figure 6. Universal DEGs in chemosensory organs. (A) A Venn diagram for DEGs in the four chemosensory organs. Numbers indicate the counts of DEGs in each organ. Chemosensory genes with differential expression in multiple organs were highlighted with arrows. (B) An UpSet plot for the same dataset visualizing the overlaps of DEGs across organs.

Discussion

Chemosensory evolution in insects provides an excellent platform to investigate the genetic architecture of the evolution of behavioral traits. Here and in this context, we used an evolutionarily young mustard specialist drosophilid fly *Scaptomyza flava* as a model to illuminate key genetic changes underlying chemosensory evolution through comparative genomics approaches with its microbe-feeding relatives.

Previous works identified six gene families serving a central role in insect chemoreception, including olfactory receptor (*Or*), gustatory receptor (*Gr*), ionotropic receptor (*Ir*), odorant-binding protein (*Obp*), pickpocket (*Ppk*), and transient receptor potential (*Trp*) gene families as key genes driving chemosensory evolution in *S. flava* (Goldman-Huertas et al. 2015; Matsunaga et al. 2022; Peláez et al. 2022) and in other insect species (Robertson 2019).

Additionally in this chapter, we conducted a series of analyses shedding light on a missing component in previous studies – the evolution of gene expression of the chemosensory gene families – based on comparisons between *S. flava* and its microbe-feeding relative, *S. pallida*. To mitigate statistical issues inherent to interspecific gene expression analyses (Price et al. 2022), we conducted multiple pipelines of DE analysis, in addition to the investigation of expression divergence between orthologous gene pairs of chemosensory families. These analyses revealed candidate genes or key gene families likely associated with the adaptation to the mustard plants in *S. flava* via the transcriptional evolution.

Expression divergence across chemosensory gene families. Our DEG analysis and pairwise expression divergence analysis using Euclidean distance and $\Delta\tau$ revealed higher gene expression divergences in the *Obp* and *Ppk* families (Figure 4A, 4C; Table 3). In a previous analysis (Peláez et al. 2023), these two families were characterized by lower gene turnover rates (i.e., frequencies of gene duplications and gene losses) compared to canonical chemoreceptor families (e.g., *Ors*, *Grs*, and *Irs*) across the *Scaptomyza* phylogeny. It is possible that these two families were under evolutionary constraint with respect to gene turnover, and instead contributed to chemosensory evolution via the modification of gene expression. Unfortunately, many candidate genes from *Obps* and *Ppks* remain functionally elusive and it is difficult to associate how their gene expression evolution is tied to the niche shift in the *S. flava* lineage.

Expression divergence in recently duplicated genes. Duplicated chemosensory genes in *Drosophila* tend to evolve faster than single-copy genes (Gardiner et al. 2008; Almeida et al. 2014), due to the relaxation of purifying selection and the subsequent positive selection in some cases. We addressed whether and to what extent this pattern holds

true for the evolution of gene expression patterns. Our analyses revealed that the degrees of expression divergence between orthologues were not significantly different between single-copy and duplicated genes (Fig. 4B, 4D). This result was consistent with the findings from a phylogenetically informed transcriptome analyses of six *Drosophila* species, wherein only 2 out of 85 recently duplicated chemosensory genes gained a gene expression in a new tissue (Bontonou et al. 2024). These patterns may indicate that gene duplication events do not necessarily facilitate changes in the evolution of gene expression in chemosensory genes. In other words, genes deviating from this pattern might be good candidates contributing to the sensory evolution in the genus *Scaptomyza*. One such example is *SflaOr98aL4/L5*. These two copies were part of the *Or98aL* clade, which experienced several duplication events in the herbivorous *Scaptomyza* lineage (Peláez et al. 2023). While expression of *S. pallida* orthologue (*SpalOr98aL3*) was limited to antennae, both of *SflaOr98aL4* and L5 copies expanded into maxillary palps (Fig. 3A). This gain of gene expression might have helped the evolution of novel behavior via the maxillary palp, such as long-range attraction to or enhanced feeding by mustard-derived chemicals (Shiraiwa 2008; Dweck et al. 2016; Oh et al. 2021).

Functionally annotated genes and their potential contributions to the adaptation to mustards. Several genes highlighted in our gene expression analysis have known functions/ligands identified in previous studies on *Scaptomyza* chemosensation and detoxification. Here, we discuss a handful of promising candidates and potential scenarios of how they are associated with the life history evolution in *S. flava*.

Or67b encodes a general odorant receptor detecting several odors/compounds such as apple cider vinegar and green leaf volatiles (GLVs) in *D. melanogaster* (Münch and Galizia 2016), and this response profile is largely conserved at least in *S. pallida* (Matsunaga et al. 2022). In the *S. flava* lineage, however, *Or67b* underwent gene duplication, resulting in three paralogous copies. The drastic shift in the chemical tuning of the OR67b receptors in *S. flava* toward volatile isothiocyanates (ITCs), key compounds characterizing chemical defenses of the Brassicales, likely mediates odor-driven attraction to their host plants (Matsunaga et al. 2022). Our analyses further revealed that *SpalOr67b* is widely expressed in antenna, tarsus, and slightly in maxillary palp, whereas the three *SflaOr67b* copies possess strong antenna-specific expression patterns (although *SflaOr67b3* is also slightly expressed in maxillary palps) (Fig. 3A). This trend of antenna-specific *Or67b* expression in the *S. flava* lineage can also be traced in τ, a parameter of the tissue specificity of gene expression (*SpalOr67b*: 0.563, *SflaOr67b1*: 0.809, *SflaOr67b2*: 0.757, *SflaOr67b3*: 0.683). These results might indicate that the microbe-feeding ancestors of *Scaptomyza* used OR67b in the legs to sense chemical cues. However, as *S. flava* transitioned to the Brassicales and the OR67b

ligand specificity shifted toward ITCs (Matsunaga et al. 2022), the tarsal expression of *Or67b* has been lost because it is no longer ecologically relevant.

The *Ir7c* gene also encodes a chemoreceptor protein, together with co-receptors IR25b and IR76b, to detect and mediate aversion to high concentrations of monovalent salt (e.g., NaCl) (McDowell et al. 2022). While *SpalIr7c* is largely tarsus-specific in expression, expression of *SflaIr7c* is elevated in olfactory organs (Fig. 3C). Interestingly, *Ir7c* is robustly expressed even in maxillary palp that lacks *Ir76b* expression (Fig. 3C), despite the fact that the loss of the IR76b co-receptor entirely eliminated the salt sensitivity in *D. melanogaster* (McDowell et al. 2022). Therefore, this expression change might indicate that IR7c have acquired novel functions, possibly in olfaction, in combination with other co-receptors like IR8a or IR25b.

Ppk23, *Ppk25*, and *Ppk29* are well known as a set of genes involved in sex pheromone detection in male *D. melanogaster* at the onset of courtship behavior (Liu et al. 2012; Thistle et al. 2012; Toda et al. 2012). Male *D. melanogaster* have two gustatory neurons specialized in pheromone sensing, called the F cell and the M cell, which mediate the detection of female- and male-specific hydrocarbons, respectively. All *Ppk* genes mentioned are expressed in the F cell, while at least *Ppk23* and *Ppk29* are expressed in the M cell (Liu et al. 2012; Lu et al. 2012; Starostina et al. 2012; Thistle et al. 2012; Toda et al. 2012). Structures and functions in vertebrate homologs of PPK channels suggested that proteins in this family tend to form a heterotrimeric ion channel (Jasti et al. 2007; Harris et al. 2008), which supports the idea that PPK23/25/29 together form a single ion channel in these pheromone-sensing neurons. However, whether the PPK23/25/29 channel directly senses sex pheromones remains controversial. As an alternative hypothesis, it is also possible that pheromones were detected by yet unknown chemoreceptors and the PPK channels only amplify the excitability of the neurons, as seen in PPK25 channel in *Or47b*-positive neurons in *D. melanogaster* (Ng et al. 2019). Our analyses in *S. pallida* revealed that all three *Ppk* genes are primarily expressed in tarsus, consistent with the knowledge in *D. melanogaster*, and *Ppk25* also shows some expression in antenna and maxillary palp (Fig 3E). On the other hand, *S. flava* drastically expanded the expression of the three *Ppk* genes including: 1) increased labellar expression in all genes, 2) expression of *Ppk29* in olfactory organs, and 3) upregulation of *Ppk23* in tarsus (Fig. 3F). Potential phenotypic changes aided by these gene expression evolution would include the involvement of labellar gustatory neurons in courtship rituals, the detection of novel food-related compounds via PPK channels, or the signal amplification of pre-existing sensory neurons. However, it is still difficult to disentangle the consequences of these expression changes, because *Ppks* play several different roles in many locations thus their function remains elusive even in *D. melanogaster*.

A total of 60 genes were differentially expressed in all comparisons, which we called “highly universal DEGs” and did not include any chemosensory genes (Fig. 6A). This result makes sense considering the difference in the repertoires of chemosensory gene expression between the four organs (Fig. 3A-E). Some genes might be associated with the evolution of herbivory in *S. flava*, such as *GstE2* (Fig. 6A). *GstE2* is a member of the large Glutathione-*S*-transferase (*Gst*) gene family encoding enzymes involved in the metabolism of environmental toxins. In fact, *S. flava* experienced the functional evolution and the gene duplication in one of the *Gst* subfamily, which enhanced the efficiency of detoxification of Brassicale-derived toxins, isothiocyanates (ITCs) (Gloss et al. 2014). Interestingly, *GstE2* gene was dramatically downregulated in *S. flava* in all eight DE comparisons, likely indicating that substrates processed by GSTE2 enzyme were less abundant in the ecological niche of *S. flava*. Contribution of detoxification genes in the sensory systems for ligand degradation has been discussed elsewhere (e.g., (Baldwin et al. 2021)). Gene expression of those detoxification gene families will need more investigation in the future.

Conclusion

Gene expression evolution of the chemosensory gene families is one essential mechanism driving behavioral evolution in insects. Here, using the young mustard-specialist drosophilid fly *Scaptomyza flava* as a model system, we explored how and what types of chemosensory genes could influence the evolution of novel behavior, in comparison with a microbe-feeding relative, *Scaptomyza pallida*. Our multiple lines of analyses revealed larger evolutionary patterns, such as the larger expression divergence in the *Obp* and *Ppk* families, and highlighted several promising candidate genes – including the ones encoding chemoreceptors for salt, sex pheromone, and mustard-derived toxins – which might have contributed to the adaptation to the mustard plants. This work laid out a foundation for the functional validation in the future, which will accelerate our understanding of the molecular mechanisms of sensory evolution.

Chapter 3. Alternative splicing of *TrpA1* gene contributes to the evolution of novel gustatory preference in *Scaptomyza flava*

Experiments and analyses in this chapter were also contributed by the following co-authors: Julianne N. Pelaez, Diler Haji, Teruyuki Matsunaga, Ashleigh Suzu Takemoto, Benjamin Goldman-Huertas, Srivarsha Rajshekar, Gary Karpen (University of California, Berkeley), Shigeru Saito, Claire T. Saito (Nagahama Institute of Bio-Sciences, Japan), Makoto Tominaga, Takaaki Sokabe (National Institute of Physiological Sciences, Japan), Kentaro M. Tanaka, Aya Takahashi (Tokyo Metropolitan University, Japan), and Noah K. Whiteman (University of California, Berkeley).

Abstract

The genetic changes underlying the evolution of novel behavioral traits remain elusive. We addressed this question by investigating the molecular basis behind the gustatory evolution in an exclusive mustard-attacking fly, *Scaptomyza flava*, and its two related microbe-feeding species, *S. pallida* and *Drosophila melanogaster*. We focused on a chemosensory ion channel, TRPA1, a primary sensor for pain-inducing chemicals including mustard-derived toxin, isothiocyanates (ITCs). Our integrative analyses revealed weaker behavioral aversion toward ITCs, desensitized activity to chemical stimuli by TRPA1 channels, and enrichment of chemically less-active TRPA1 splice isoforms in *S. flava*, all of which together suggested that both protein-level and expression-level mechanisms drove the gustatory evolution toward ITCs in a mustard-specialist. However, the chemical sensitivity of TRPA1 was even lower in the distantly related *D. melanogaster,* suggesting that TRPA1 underwent a complicated evolutionary trajectory across the drosophilid flies.

Introduction

Behavioral traits are often coded by complex genetic architecture, thus it remains elusive at the molecular level how behavioral variation arises or behavior diverges across lineages (Bendesky and Bargmann 2011). Insect chemosensation provides appealing windows for addressing such evolutionary questions, because of their simplicity and modularity of functional units, where the characteristics of a single receptor often defines the behavioral response to a given stimulus (Montell 2021). The genus *Drosophila* has been serving a central role in this field. Accelerated by a wealth of genomic resource and genetic tools, comparative analyses including several ecological specialists, such as *D. sechellia* (McBride 2007; Prieto-Godino et al. 2017; Auer et al. 2020), *D. erecta* (McBride and Arguello 2007), *D. suzukii (Karageorgi et al. 2017; Dweck et*

al. 2021), and *D. mojavensis* (Diaz et al. 2018; Khallaf et al. 2020), have illuminated the genetic basis of chemosensory evolution. The latest to this lineup is *Scaptomyza flava*, a specialist herbivore of the mustard plants and relatives that produced glucosinolates (in the order Brassicales). Unlike their microbe-feeding relatives, both larvae and female adults of *S. flava* actively feed on leaves of their Brassicales host plants that are chemically distinct from microbial diets (Whiteman et al. 2011; Aguilar et al. 2024), suggesting that natural selection posed by plant chemicals likely drove the chemosensory evolution in this species. Previous analyses estimated the origin of herbivory in the *S. flava* lineage at ca. 10-15 million years ago (Lapoint et al. 2013; Peláez et al. 2022), and such a young dietary transition is suitable for comparative analyses to track down the genetic changes accompanying the diet shift. These characteristics of *S. flava* make it an appealing target for the study of the molecular basis of chemosensory evolution too.

The major defense system equipped by mustard plants is a group of tissue-damaging toxins called isothiocyanates (ITCs) or mustard oils. Brassicales normally possess lineage-specific secondary metabolites, glucosinolates (GLSs), which themselves are not toxic. When the plant tissues are mechanically damaged by herbivores, however, GLSs are hydrolyzed by an enzyme myrosinase and converted into ITCs, a mechanism known as “mustard oil bomb”, which are highly toxic and induce robust escaping behaviors in generalist insects including the fruit fly *D. melanogaster* (Lichtenstein et al. 1964; Kang et al. 2010). Acquisition of this chemical defense system conferred the protection against herbivores and further stimulated the diversification of the family Brassicaceae (Edger et al. 2015). Coevolutionary arms race has also shaped the counter-defense system against ITCs on the herbivore side. For instance, a group of mustard-specializing butterflies, such as the cabbage white *Pieris rapae*, can suppress the formation of ITCs by the action of gut-expressed proteins called nitrile specifier proteins (NSPs), which convert the hydrolysis products of GSLs into less toxic nitriles and allow the butterflies to feed on Brassicales without having to deal with a high dosage of ITCs (Wittstock et al. 2004; Wheat et al. 2007; Okamura et al. 2022). By contrast, *S. flava* does not have such effective detoxification mechanisms. Although they can mitigate mustard oils by an ancestral mechanism called mercapturic acid pathway (Gloss et al. 2014), this is a less efficient detoxification mechanism, which indicates that *S. flava* still encounters ITCs while feeding on the mustard plants. Therefore, suppression of behavioral aversion against mustard oils appears to be essential for *S. flava* to evolve into an active Brassicales feeder.

One candidate gene underlying this putative sensory change is called *TrpA1*, part of the *Trp* (*Transient receptor potential*) channel gene family in insects. TRPA1 is a conserved polymodal cation channel that can be activated by several kinds of pain-inducing stimuli, such as heat, UV, reactive oxygen species, and chemicals

including ITCs (Jordt et al. 2004; Talavera et al. 2020). Most TRPA1 agonists including ITCs are electrophiles that form covalent modifications with several cysteine residues in the cytoplasmic domain of TRPA1 and thereby change the conformation of the entire channel (Hinman et al. 2006; Macpherson et al. 2007), a mechanism conserved from vertebrates to flies and planarians (Kang et al. 2010; Arenas et al. 2017). In adult *D. melanogaster*, TRPA1 serving chemical detection is expressed in bitter-sensing neurons of the mouthpart (labellum and pharynx) and the legs (tarsi) that triggers feeding aversion upon activation (Kang et al. 2010; Leung et al. 2020). Genetic knockout of *TrpA1* entirely disrupts the aversion against electrophiles, implying TRPA1 channel is the primary gustatory sensor for drosophilids to detect dietary electrophiles (Kang et al. 2010). These facts suggest that the gustatory preference to ITCs in flies could potentially evolve by the molecular evolution of the TRPA1 channel.

There are a few actual cases where TRP channels seemingly drove sensory evolution in other animals. For example, one subspecies of the African mole-rats called Highveld mole-rat (*Cryptomys hottentotus pretoriae*) is entirely insensitive to a pain-inducing chemical allyl isothiocyanate (AITC, commonly known as wasabi) possibly as an adaptation to ant-derived toxins in their diets (Eigenbrod et al. 2019). The lack of pain sensation in this species is correlated with the reduction of chemical sensitivity of TRPA1 by the substitution of one of the key cysteine residues, providing a simple molecular explanation for the behavioral change (Eigenbrod et al. 2019). Variation in heat activation was also observed among reptiles and amphibians which corresponds to species-specific thermal niches (Gracheva et al. 2010; Cordero-Morales et al. 2011; Akashi et al. 2018; Saito et al. 2022), although exact amino acid substitutions causing the functional shifts were not identified in most cases.

A slightly more complex example came from TRPV1 channel, another vertebrate heat sensor, in the vampire bat (*Desmodus rotundus*) that can sense infrared signals with their pit organs. In a comparison with its non-infrared-sensing relative, a study found that TRPV1 in these bats produced two splicing isoforms with different thermal sensitivity, and that the vampire bat trigeminal neurons (running near pit organs) were highly enriched by a TRPV1 isoform with lower activation temperature, which likely enables to detect slight temperature changes by infrared (Gracheva et al. 2011). In fact, *D. melanogaster* TRPA1 produces five splice variants with salient functional differences among them (Kang et al. 2012; Gu et al. 2019) that could potentially modulate neural activity via alternative splicing.

Taken together, these examples highlighted two molecular mechanisms that could potentially mediate sensory evolution in *S. flava* via TRPA1: 1) desensitization of the channel, and 2) enrichment of isoforms with weaker chemical responses. Furthermore, studies from other sensory receptors hinted at other possibilities, such as: 3) up- or down-regulation of TRPA1 expression in the gustatory organs, or 4)

misexpression of TRPA1 in sugar-sensing neurons. If TRPA1 is involved in the gustatory evolution in the mustard specialist lineage, we expect to observe some of these molecular changes in *S. flava*.

In this chapter, we hypothesized that the molecular evolution of TRPA1 channel facilitated the specialization to Brassicales host plants in the *S. flava* lineage by weakening the aversion to ITCs, and thoroughly investigated the characteristics of TRPA1 across the phylogeny and their association with the dietary evolution. To address this hypothesis, we took a comparative approach between three species from different phylogenetic positions: the representative generalist *D. melanogaster*, the Brassicales specialist *S. flava*, and its closely related microbe-feeder *S. pallida*. Between the three focal species, we compared behavioral responses against ITCs, gene-level and isoform-level expression of *TrpA1* gene, and physiological responses of TRPA1 channel isoforms against electrophiles. We also performed another set of behavior assay to see whether phenotypic differences between species can be recapitulated by transgene expression of *TrpA1* in bitter-sensing neurons. Our results suggested that TRPA1 traced rather complicated evolutionary paths across the *Drosophila* phylogeny, but the overall desensitization of and the skewed expression of less active isoforms are indeed the two key events underlying the acquisition of the taste for mustards in *S. flava*.

Materials and methods

Fly strains. The *Scaptomyza flava* laboratory colony was originally established from >150 wild larvae collected from Dover, NH (USA) in 2008. The colony was maintained in cages containing *Arabidopsis thaliana* Col-0 strain as larval and adult food source at 20-21°C and 60% humidity in a walk-in environmental room on a 12L:12D cycle. The *Scaptomyza pallida* lab colony was an isofemale line established from a wild female fly collected at the University of California, Berkeley campus (CA, USA) in 2017. The colony was reared on organic spinach (purchased at Trader Joe's, CA, USA) placed on top of regular cornmeal media purchased from UC-Berkeley Fly Food Facility. Other rearing conditions were identical to those of *S. flava* colony. All wild-type and transgenic *Drosophila melanogaster* lines were obtained from Bloomington Drosophila Stock Center, except for the ones newly created in this study, and reared on cornmeal media from UC-Berkeley Fly Food Facility.

Behavioral assay (Wild-type). In order to characterize behavioral preferences to electrophiles in the three focal species, we conducted the proboscis extension response (PER) assay using allyl isothiocyanate (AITC) mixed in sucrose solution. PER is a commonly used method in *Drosophila* to measure feeding preference to a given chemical by monitoring the frequency of proboscis extension after the presentation of test chemicals (Shiraiwa and Carlson 2007). Our protocol for PER assay was based on

the one from (Kang et al. 2010) with several modifications. Adult flies were collected from their larval media soon after the emergence and transferred to plastic vials with 10% honey water (pure organic honey purchased at Trader Joe's, CA) to account for larval food differences. We did not separate male and female flies during this period, thus we assume that all female flies were already mated at the time of experiments. 2-9 day old flies were starved for 24 hours in a humidified vial prior to the assay. Flies were then anesthetized with CO_2, mounted on a glass slide dorsal side down using nail polish, and allowed to recover in a humidified chamber for at least 2 hours. Flies that did not show active body movement after the recovery period were removed from the assay. During the PER assay, all flies were first satiated with water, and each fly was offered a single type of tastant (i.e., pure sucrose water (300mM) or its mixture with a given concentration of allyl isothiocyanate (AITC)) by touching their forelegs with test chemicals as a small droplet from a P200 pipette tip. We picked AITC as a representative ITC for this and subsequent experiments, because it is the most commonly used ITC in previous laboratory-based studies using *Drosophila*. Each presentation lasted up to 10 seconds, and the response of the fly to the tastant within the 10-second window was scored based on our criteria: 1pt - Extended proboscis and maintained drinking the droplet >2 seconds; 0.5pt - Extended proboscis but failed to maintain contact to the droplet, or only showed brief contact; 0pt - Did not show proboscis extension and failed to contact the droplet. The presentation was repeated five times per fly (with a 5-10 minute interval between presentations), such that each fly earned a total score ranging from 0 to 5. We performed PER assay at four different AITC concentrations (i.e., 0, 2, 5, 10mM AITC in a 300mM sucrose solution), assuming the biological ITC concentration for the Brassicales plants were somewhere between 1mM to 5mM. The difference in PER scores across the AITC dosage series was tested for each group using pairwise Wilcon's rank-sum test.

Starvation tolerance assay. Adult flies were collected upon emergence and reared on 10% honey water. 8-10 flies (2-5 days old) of each species/sex were grouped and transferred to a plastic vial containing 1% agar, and incubated under a laboratory condition (20-21°C, 60% humidity, 12:12 L:D cycle). We then monitored the number of dead/live flies every 12 hours until all individuals were dead. We tested 5-6 replicates for each species/sex group. Survival rates differences between species were analyzed for each sex by pairwise log-rank test using survminer package v0.4.9 (Kassambara et al. 2017) in R v4.2.1.

RNA-sequencing. We performed bulk RNA-sequencing to confirm *TrpA1* expression is not lost in *S. flava*, to compare expression levels between species, and to characterize *TrpA1* isoform cDNA sequences in *Scaptomyza*. Total RNA extraction from *S. flava* and

S. pallida adult females, RNA-sequencing, and initial filtering of RNA-seq reads were performed as described in Chapter 2. To roughly predict mRNA structures of *Scaptomyza TrpA1*, we first conducted *de novo* assembly of transcript contigs using Trinity v2.5.1 (Grabherr et al. 2011) and conducted TBLASTN searches (v2.10.1) against the assembled contigs of *S. flava* and *S. pallida* using *D. melanogaster* TRPA1-C protein sequence as a query. Exonic structures of *S. flava* and *S. pallida TrpA1s* were manually inspected after sequence alignment using MAFFT v7.273 (Katoh and Standley 2013). Differential expression analysis between species was conducted as described in Chapter 2. We also downloaded previously published RNA-seq data from female *D. melanogaster* labellar to compare gene expression with *S. flava* in the same manner.

Fluorescent *in situ* hybridization. We conducted RNA FISH to examine the types of gustatory neurons expressing *TrpA1* in the fly labellum. HCR (hybridization chain reaction) RNA FISH kit was ordered from Molecular Instruments (CA, USA). Custom mRNA probes for *D. melanogaster* and *S. flava TrpA1* (and respective *Gr66a* as a marker for bitter-sensing neurons) were designed and produced at Molecular Instruments (CA, USA). *TrpA1* probes were targeted spanning exons 5-11 so they can bind to all splice isoforms equally.

Female *D. melanogaster* and *S. flava* between 3-10 days old were anesthetized with CO_2. Their mouthparts were hand-dissected with fine forceps and fixed in a 2mL fixative solution (4% paraformaldehyde in 1X PBS, with 0.1% Triton X-100) at 4°C on a nutator for 24 hours. Following the fixation, samples were washed twice in 2mL 1X PBS + 3% Triton X-100, once in 2mL 1X PBT (i.e., PBS + 0.1% Triton X-100), and four times in 1mL 1X PBT. Hybridization and amplification steps were mostly identical to the official protocol for *D. melanogaster* embryos from Molecular Instruments, except for the higher probe/hairpin concentrations (Bontonou et al. 2024): Probe solutions were created by adding 5µL of experimental (*TrpA1*) and control (*Gr66a*) probes to 300µL probe hybridization buffer. Likewise, hairpin solutions were prepared by adding 10µL of each 3µM hairpin stock to 300µL amplification solution.

Amplicon sequencing. We next aimed to quantify the proportions of expressed *TrpA1* isoforms in fly labella and tarsi to investigate whether those ratios are associated with the dietary adaptation of the flies. Due to the nature of *Scaptomyza/Drosophila TrpA1* isoforms where exons defining isoform types are distantly located (i.e., ~2kb on mRNA), common gene expression analyses such as short-read sequencing or quantitative PCR do not accurately capture the proportions of *TrpA1* splicing variants. Therefore, we employed amplicon sequencing targeting the entire *TrpA1* cDNA with long-read sequencers from Oxford Nanopore Technologies (ONT), assuming the ratio of *TrpA1* isoforms was still conserved after PCR amplification.

Total RNA was extracted from independent groups of fly labella or tarsi (all six legs pooled) where each replicate contained roughly 40-50 individuals. 100ng total labellar RNA or 50ng tarsal RNA was transcribed into cDNA using SuperScript IV (Invitrogen, USA), which was later used as templates for first-round PCR for *TrpA1* transcripts using Phusion high-fidelity DNA polymerase (New England Biolab). PCR primers were designed for the common 5'UTR and the last exon such that the entire pool of *TrpA1* isoforms can be amplified in a single reaction. Furthermore, the primers were concatenated to common "adapter" sequences acting as annealing sites for barcoding primers in the second-round PCR. Successful *TrpA1* amplicons were isolated from 1% agarose gel after electrophoresis and purified using Monarch DNA gel extraction kit (NEB). 40ng *TrpA1* amplicons were then used as templates for barcoding PCR using LongAmp Taq 2X Master Mix (NEB) and PCR Barcoding Expansion (EXP-PBC001, ONT). Barcoded *TrpA1* fragments were cleaned with QIAquick PCR Purification Kit (Qiagen) and quantified with Qubit dsDNA BR Assay Kit (ThermoFisher). Subsequently, multiple *TrpA1* amplicons were pooled together in a single tube for size selection using X0.6 SPRI beads. Selected fragments were used for library preparation using Ligation Sequencing Kit (SQK-LSK109, ONT) and sequenced on Flongle Flow Cell (R9.4.1) with MinION (ONT). After obtaining FASTQ output files, we performed BLASTN searches against the outputs using pre-characterized *TrpA1* sequences as queries to figure out the isoform identity of each single read. To take into account the high error rates of ONT sequencers, we filtered out the reads where the lengths are outside the range between 3.8 and 4.5 kb or the sequence similarity to the best hit query was less than 85%. Passed reads were categorized as their best-hit isoform type, and the proportions of isoforms were calculated using a manual Python script. The relative proportion of each *TrpA1* isoform type between species/sexes were analyzed using R v4.2.1 using generalized linear model assuming gaussian distribution.

Oocyte electrophysiology. We performed two-electrode voltage clamp recordings using *Xenopus* oocytes to characterize chemical sensitivities of detected TRPA1 isoforms. Ten *Scaptomyza*/*Drosophila TrpA1* isoform CDSs were subcloned into a *Xenopus* expression vector pOX+ (citation) using NEBuilder HiFi DNA Assembly Master Mix (NEB). Original *TrpA1* sequences were obtained from various sources summarized in Table 1. cDNA synthesis was conducted with 40-80ng starting total RNA using Superscript IV (Invitrogen). Note that here we denote *TrpA1* orthologues with initials of each species. *D. melanogaster TrpA1* is commonly called "*dTrpA1*", but instead we called it "*DmTrpA1*" for consistency. pOX(+)-*TrpA1* constructs were linearized with a restriction enzyme RsrII and used as templates for cRNA synthesis using mMessage mMachine SP6 Transcription kit (ThermoFisher).

Isoform	Original source	Amplification
SfTrpA1-CL	Artificial synthesis (Genscript, CA, USA)	Competent cells
SfTrpA1-DL	Artificial synthesis (Genscript, CA, USA)	Competent cells
SfTrpA1-CS	*S. flava* labellar cDNA	PCR (Q5 high-fidelity DNA polymerase, NEB)
SfTrpA1-DS	*S. flava* labellar cDNA	PCR (Q5 high-fidelity DNA polymerase, NEB)
SpTrpA1-CL	*S. pallida* labellar cDNA	PCR (Q5 high-fidelity DNA polymerase, NEB)
SpTrpA1-DL	*S. pallida* tarsal cDNA	PCR (Q5 high-fidelity DNA polymerase, NEB)
SpTrpA1-EL	*S. pallida* labellar cDNA	PCR (Q5 high-fidelity DNA polymerase, NEB)
SpTrpA1-CS	*S. pallida* labellar cDNA	PCR (Q5 high-fidelity DNA polymerase, NEB)
DmTrpA1-C	*D. melanogaster* larval whole-body cDNA	PCR (Q5 high-fidelity DNA polymerase, NEB)
DmTrpA1-D	A gift plasmid from Marco Gallio's lab	Competent cells

Table 1. Sources of subcloned *TrpA1* isoform sequences.

Oocytes were collected from adult *Xenopus laevis*, processed with Collagenase A (Roche) to remove follicular membranes, and incubated in MBSH-PH solution until cRNA injection. 50nL cRNA of a given TRPA1 isoform (or a mixture of several isoforms) was injected into each oocyte. Two-electrode voltage clamp (TEVC) recordings were performed 2-3 days after the cRNA injection using OC-725C amplifier (Warner Instruments) and Digidata 1440 (Axon Instruments). All recordings were performed at -60mV. Taking advantage of the polymodal nature of fly TRPA1, we stimulated the channels twice in the course of a single recording: oocytes were first heated (42°C) for 30 seconds, washed for 120 seconds, and then challenged with a given dosage of AITC for 90 seconds. This design allowed us to normalize a chemically induced raw current by the amplitude of heat-induced current at each recording (assuming heat current amplitudes reflect the abundance of TRPA1 channels in oocytes). Recordings were performed at a series of concentrations ranging from 0.001mM to 3mM. We used heat-normalized current values to estimate dose-response relationships and EC_{50} for each TRPA1 isoform using R package drc v3.0-1 (Ritz et al. 2015). Interspecific size differences in normalized electrophile currents at each concentration, as well as heat-induced raw currents were analyzed with Kruskal-Wallis test, followed by Dunn's test with Benjamini-Hochberg correction using R v4.2.1.

Generating transgenic flies. We further performed genetic rescue experiments in *TrpA1*-mutant *D. melanogaster* lines to test whether the ectopic expression of

Scaptomyza TRPA1 sufficiently recapitulates the behavioral variation of wild-type species. Three *TrpA1* isoforms (*SfTrpA1-CL*, *SpTrpA1-CL*, *DmTrpA1-C*) were subcloned into 5xUAS-attP vectors for fly transgenesis. These *UAS-TrpA1* cassettes were inserted into *D. melanogaster* genome targeting at attP40 site on chromosome 2 using PhiC31 integrase. Transgenesis of *D. melanogaster* embryos and the initial transformant screening were performed at Genetivision (Huston, TX, USA), which resulted in three independent *UAS-TrpA1* lines. Other mutant strains were obtained from Bloomington *Drosophila* Stock Center (#28801: *pGr66a-GAL4;Gr93a*^{3}, #26504: *w***;TrpA1*^{1}, #7198 *w***;CyO/Kr*$^{If\text{-}1}$*;TM3,Ser*^{1}*/D*^{1}). All mutant lines were initially backcrossed to *w*^{1118} for five generations to homogenize their genetic background. *pGr66a-GAL4* (after omitting the *Gr93a*^{3} allele) and all *UAS-TrpA1* lines were then crossed to *TrpA1*^{1} (= *TrpA1*-null mutant) lines so all flies were null mutants for the endogenous *TrpA1* allele. Each *UAS-TrpA1;TrpA1*^{1} line was then crossed either to *pGr66a-GAL4;TrpA1*^{1} line (test) or to *+/+;TrpA1*^{1} line (control) to obtain the genotypes we used in experiments.

Behavioral assay (Transgenic lines). Procedures of the PER assay for transgenic flies were identical to those for wild-type flies except for these modifications: (1) Adult flies were collected soon after the emergence and kept on fresh cornmeal media until the 24-hour starvation period, (2) 3-6 day old female flies were used for the assay (assuming all flies were mated by this period), (3) after the starvation period, flies were inserted into P200 plastic tips such that only the head and forelegs were exposed and allowed to access tastants. As test substances, three concentrations (0, 5, 10mM) of N-methylmaleimide (NMM) were prepared in 100mM sucrose solution containing 0.5% DMSO. Here, we switched from AITC that could interfere with the endogenous olfactory pathway of flies, and instead used another TRPA1 activator, NMM, so we can evaluate the effect of TRPA1s in gustatory behavior. Each fly was tested with a single tastant five times, in which each presentation lasted up to 5 seconds. All assays were performed between 3 to 6 pm. PER scores were calculated in the same criteria as those in wild-type PER, and differences in the distribution of total PER scores were statistically tested within genotypes using Wilcoxon's rank-sum test with Benjamini-Hochberg correction.

Results

Gustatory aversion from isothiocyanates is attenuated in *S. flava*. Although we anecdotally observed the lack of aversion in *S. flava* from mustard plants, their gustatory preference for ITCs have not been quantitatively investigated. We began this study by characterizing behavioral preference to an isolated electrophilic compound, allyl isothiocyanate (AITC), across *Scaptomyza*/*Drosophila* species by monitoring PER

(proboscis extension response) to tarsal/labellar stimulation. We used a given concentration of AITC mixed in 300mM sucrose solution as test stimuli to investigate the degree of feeding inhibition by the presence of AITC in the three wild-type species. PER assay showed that the two microbe-feeding species, *D. melanogaster* and *S. pallida*, clearly exhibited a stepwise reduction of feeding responses in both sexes as the concentration of AITC increased (Fig. 1A). By contrast, *S. flava* did not show feeding aversion up to 5mM AITC (Fig. 1A). Male *S. flava* significantly avoided the tastant at 10mM AITC, but female flies were not inhibited at that dosage (although the average PER score was lower compared to 0mM AITC) (Fig. 1A). These data suggest that the behavioral aversion to AITC in *S. flava* is not entirely lost, but it has been significantly attenuated relative to its microbe-feeding relatives, consistent with their mustard-dependent ecologies.

One scenario to explain the observed behavioral differences is that *S. flava* was more hungry than other species after the 24-hour starvation period prior to the PER assay, such that bitter sensation at the periphery was overridden by strong hunger-induced appetitive signals (Inagaki et al. 2014; Yapici et al. 2016). To explore this possibility, we ran a simple hunger tolerance assay to monitor the survival rate of the flies in foodless humidified vials (Li et al. 2016). We found that hunger resistance was significantly different between all three species in each sexes (Fig. 1B; Log-rank test, *p-value*<0.001 for all pairwise comparisons) and that the starvation tolerance of *S. flava* was intermediate between those of *D. melanogaster* and *S. pallida* (Fig. 1B). In both sexes, *D. melanogaster* was most susceptible to the starvation treatment. *S. flava* and *S. pallida* were comparable up to 100-150 hours, but afterwards, *S. pallida* exceeded the tolerance of *S. flava* wherein a few *S. pallida* individuals survived almost up to two weeks without any food (Fig. 1B). This difference might be reflected in the overall lower PER rates in *S. pallida* (Fig. 1B). However, the fact that aversion to AITC was clearly observed in *S. pallida* could imply that the hunger does not well explain the variation in AITC-aversion across species. Therefore, we conclude that the attenuated aversion to AITC in *S. flava* is likely attributed to their lineage-specific modification in the sensory systems.

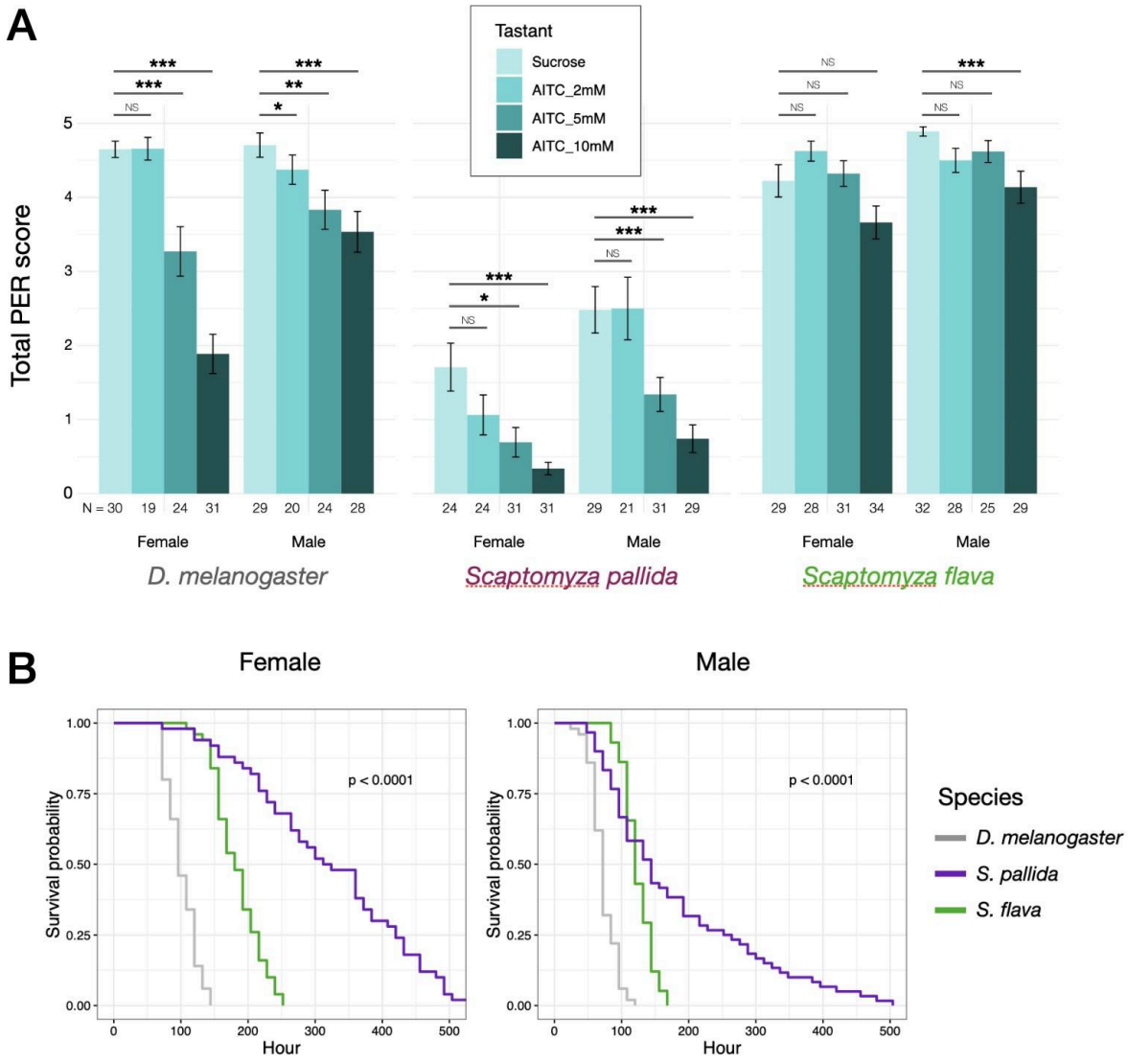


Figure 1. Electrophile sensitivities of the wild-type fly species. (A) PER responses against AITC (in 300mM sucrose) in male and female of the three drosophilid species. Reduction of PER scores within each group was tested by pairwise Wilcoxon rank-sum test (P-value: NS, not significant; *, p<0.05; **, p<0.01; ***, p<0.001). (B) Survival plots from starvation tolerance assay in (left) female and (right) male flies. Differences between species were analyzed by log-rank tests (only group-wide p-values were shown).

Spatial property of *TrpA1* expression is conserved across species. We next conducted short-read RNA-sequencing to confirm gene expression of *TrpA1* in *S. flava* and characterize their mRNA sequences. Cross-species DEG analysis revealed robust expression of *TrpA1* in labellum and tarsus in all species tested, ruling out the possibility of the loss of *TrpA1* expression in *S. flava* (Fig. 2A). The overall expression levels of

TrpA1 were not statistically different between species, although expression is almost twice as high in *D. melanogaster* as in *S. flava* (Fig. 2A). To characterize the types of neurons expressing TRPA1, we conducted HCR (hybridization chain reaction) RNA-FISH to visualize mRNA localization in the labellum of *S. flava* and *D. melanogaster*. We found in *S. flava* that signals of *TrpA1* formed neuron-like clusters and always co-localized with those of *Gr66a*, a marker of bitter-sensing gustatory neurons, consistent with the pattern observed in *D. melanogaster* (Fig. 2B) and those from a previous study (Leung et al. 2020). These results ruled out the possibility of TRPA1 misexpression and indicated that the valence of ITCs as aversive signals was still conserved even in *S. flava*.

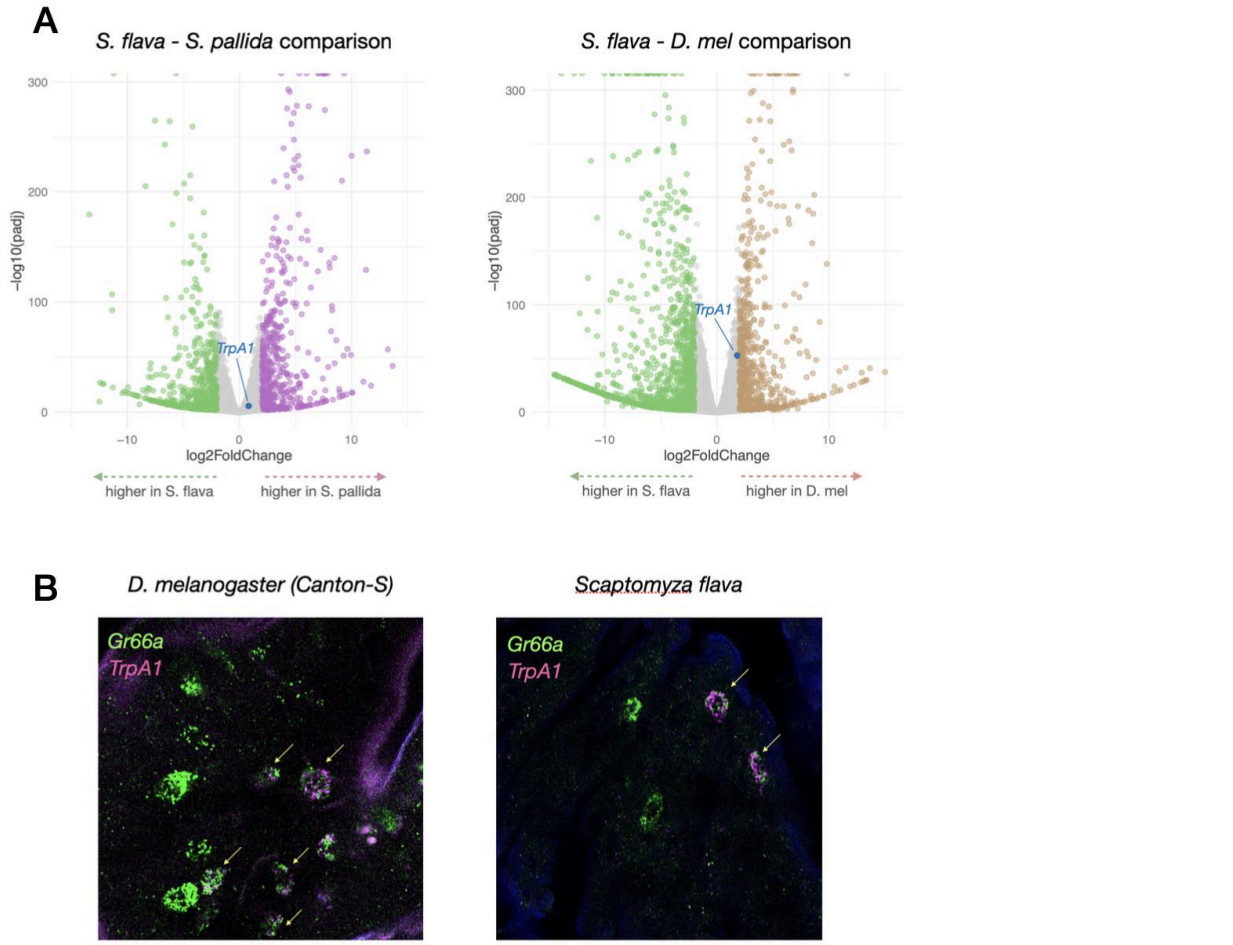
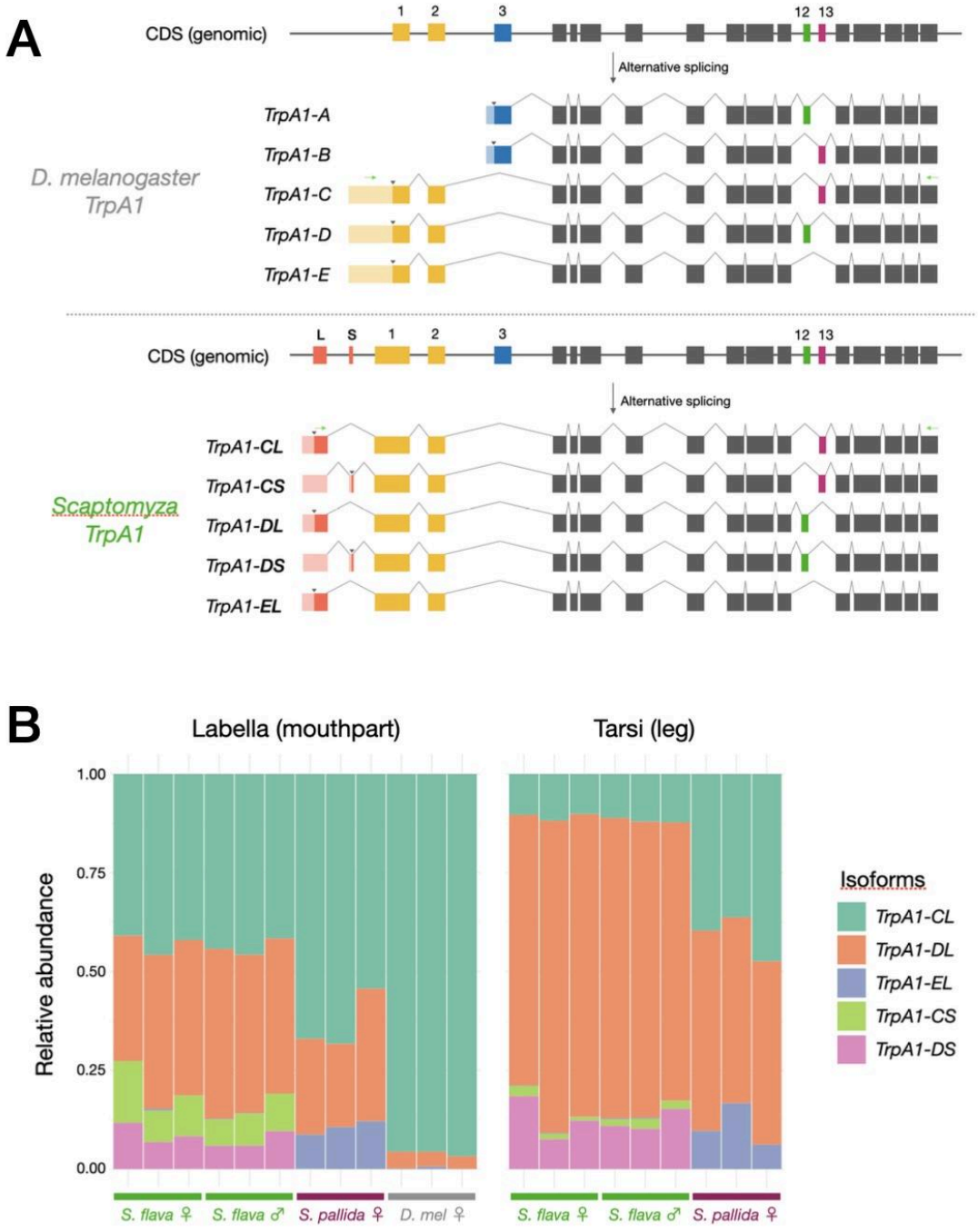


Figure 2. Gene expression of *TrpA1* in *Scaptomyza flava* compared to *S. pallida* and *D. melanogaster*. (A) Volcano plots of cross-species DEG analysis in female labella between (left) *S. flava* - *S. pallida* and (right) *S. flava* - *D. melanogaster*. Each data point shows a single gene, where colored ones are significantly differentially expressed between species. *TrpA1* (blue) was not a significant DEG in both comparisons. (B) Microscopic images of HCR RNA FISH samples: (left) female *D. melanogaster* labellum (right) female *S. flava* labellum. Clusters of green and magenta signals (pointed by yellow arrows) indicate the neuronal expression of *Gr66a* and *TrpA1*, respectively.

Compositions of *TrpA1* splicing variants are variable across species. Unlike mammalian *TRPA1*, *D. melanogaster TrpA1* gene undergoes alternative splicing upon transcription that produces five isoforms with various expression and functional properties (Zhong et al. 2012; Gu et al. 2019). We were interested in the compositions of TRPA1 channel isoforms across *Drosophila/Scaptomyza* species, because the proportion of spliced variants potentially aids sensory adaptation as in the case of TRPV1 isoforms in the infrared-sensing bat (Gracheva et al. 2011). Initial *de novo* assembly of RNA-sequencing reads revealed the presence of at least five *TrpA1* isoforms in the labella/tarsi of *S. flava* and *S. pallida* (Fig. 3A). All *Scaptomyza* isoforms contained exons 1 and 2 of *TrpA1*, showing similarities to *TrpA1-C/D/E* isoforms of *D. melanogaster* (Fig. 3A). Interestingly, *Scaptomyza TrpA1s* seemingly possessed two additional exons upstream of Exon 1, named Exon L and Exon S (meaning "Longer" or "Shorter" exons), each of which contained putative start codons (Fig. 3A). These *Scaptomyza*-specific *TrpA1* isoforms were named *TrpA1-CL/CS/DL/DS/EL*, respectively, based on the choice of exons and their homology to *D. melanogaster TrpA1* isoforms (Fig. 3A). Presence of these isoforms were further validated by subcloning and sequencing.

Quantification of isoform expression posed a challenge in the case of *Scaptomyza TrpA1*. Because exons that determine isoform identities are distantly located (i.e., exons L/S and 12/13, see Fig. 3), it was impossible to accurately parse out the expression of each single isoform by short-read RNA-seq that can only cover up to 150bp per read. To address this issue, we performed amplicon sequencing of *TrpA1* transcripts using a long-read sequencer from Oxford Nanopore Technologies (ONT) that allows sequencing up to 4Mb in a single read. Briefly, we PCR-amplified the entire repertoire of *TrpA1* transcripts from labellar or tarsal cDNA for each species (see Method section), sequenced them on Flongle flow cells using ONT MinION, and identified each sequence read using BLAST searches. This approach revealed substantial variations in the proportion of *TrpA1* isoforms across species and organs, but not between sexes (Fig. 3B). Notably, the *TrpA1-CL(C)* isoform was dominant in *D. melanogaster* and *S. pallida* labella, while it was equally abundant as *TrpA1-DL* in *S. flava* labella (Fig. 3B, 3C). Differences between tissues were also striking, wherein *TrpA1-DL* isoforms were more abundant than *TrpA1-CL* in both *S. flava* and *S. pallida* (Fig. 3B, 3C). Isoforms that contained Exon S, namely *TrpA1-CS* and *-DS*, were only detected in *S. flava* (Figure 3B). These unique proportions of *TrpA1* isoforms suggest that the differences in splice isoform compositions might be associated with the behavioral evolution across the fly lineages.


A
CDS (genomic)
1
2
3
12
13
Alternative splicing
TrpA1-A
TrpA1-B
TrpA1-C
TrpA1-D
TrpA1-E
D. melanogaster TrpA1
L
S
CDS (genomic)
Alternative splicing
TrpA1-CL
TrpA1-CS
TrpA1-DL
TrpA1-DS
TrpA1-EL
Scaptomyza TrpA1
B
Labella (mouthpart)
Tarsi (leg)
Relative abundance
1.00
0.75
0.50
0.25
0.00
S. flava ♀
S. flava ♂
S. pallida ♀
D. mel ♀
Isoforms
TrpA1-CL
TrpA1-DL
TrpA1-EL
TrpA1-CS
TrpA1-DS

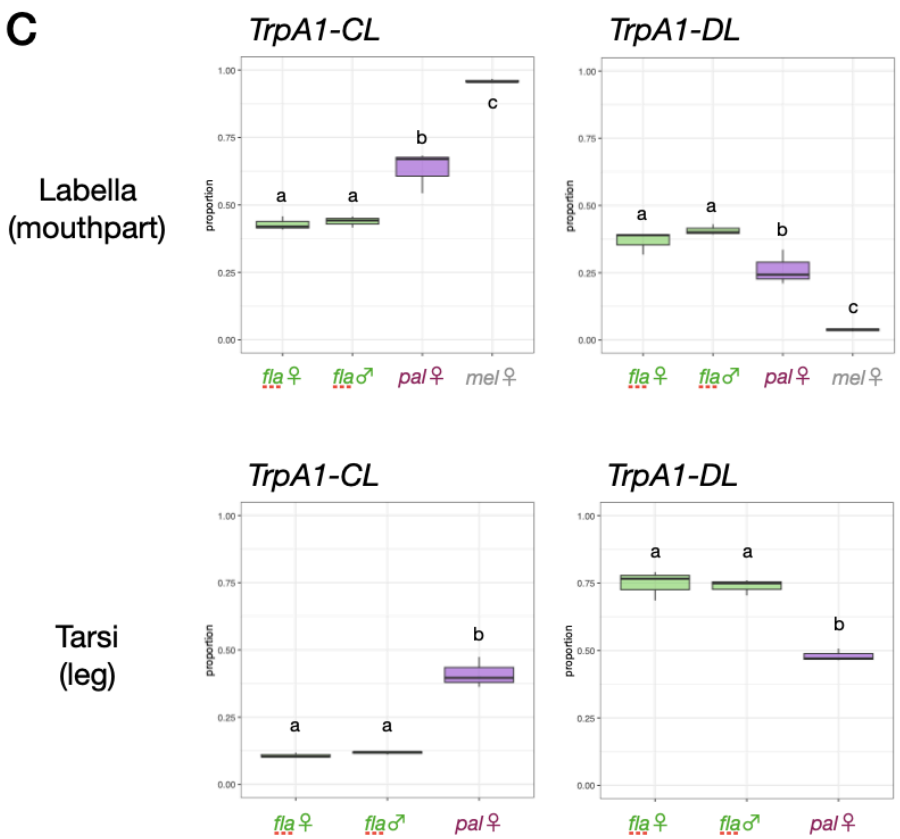


Figure 3. *Scaptomyza TrpA1* isoforms. (A) Exon-intron structures of *TrpA1* splicing variants of *D. melanogaster* and *Scaptomyza* (i.e., those identified from *S. flava* and *S. pallida* labellar cDNA). Black small arrowheads indicate putative start codons, whose upstream regions (with pale colors) show putative 5'UTR regions. Green arrows indicate the positions of PCR primers designed for amplicon sequencing. (B) Compositions of *TrpA1* isoforms from labelar and tarsal cDNA of the three focal species. Each vertical bar represents a cDNA sample derived from an independent set of 40-50 flies. Note that *D. melanogaster TrpA1-C/D/E* isoforms here were categorized as *TrpA1-CL/DL/EL*, respectively, for the sequence similarity between its 5'UTR region and Exon L of *Scaptomyza TrpA1s*. (C) Relative abundances of *TrpA1-CL* and *-DL* in labellar and tarsal cDNA (N=3 for each group). Variations between groups were modeled by GLM with Gaussian distribution and tested by Tukey's post-hoc test with Benjamini-Hochberg correction.

TRPA1 channels *in vitro* show different responses to AITC between or within species.

We next characterized functionalities of TRPA1 channel isoforms detected from the three drosophilids using two-electrode voltage clamp (TEVC) recordings from *Xenopus laevis* oocytes. As reported in a previous study (Kang et al. 2010), we also noticed that drosophilid TRPA1 channels stimulated by AITC did not immediately return to the baseline current and got desensitized to subsequent chemical stimulations, possibly due to its unique mode of channel activation (Hinman et al. 2006; Macpherson et al. 2007). This property prevented us from using a common approach wherein a raw current amplitude was normalized by another maxed-out current applied at the end of

the recording. Instead, we took a strategy in which TRPA1s were initially stimulated by heat (42d°C) and subsequently by a given concentration of AITC, such that AITC-induced currents were normalized by heat-induced currents at each recording (Fig. 4A). This approach assumed that sizes of heat-induced currents reflected the degree of expression of TRPA1 in oocytes, which in fact was supported by our dataset where the two raw current sizes were positively correlated (data not shown yet).

Using oocyte electrophysiology, we began comparing dose-responses between three drosophilid species for TRPA1-CL and -DL isoforms, which were the primary components of *Scaptomyza* TRPA1 isoforms (Fig. 3B). We found that basic responses of each TRPA1 isoform to AITC washighly conserved: For example, CL tended to be slowly activated at lower concentrations and showed acute spike-like responses at higher concentrations (Fig. 4A). By contrast, DL was mostly insensitive to AITC at lower doses while it was highly activated at higher dosages, and characterized by slower activation to the peak compared to CL (Fig. 4A). Although these modes of activation were conserved across lineages, current amplitude showed substantial variations among species. In both TRPA1 isoforms, *S. pallida* generated the largest amplitudes to AITC (Fig 4B, 4C). Responses of *S. flava* and *D. melanogaster* were comparable at lower dosages, but *S. flava* TRPA1 eventually was saturated at a larger amplitude on average (Fig. 4B, 4C). Median effective concentrations (EC_{50}) were somewhat comparable in C isoforms (DmTRPA1-C: 0.013mM, SpTRPA1-CL: 0.034mM, SfTRPA1-CL: 0.028mM), while the variation was larger in D isoforms (DmTRPA1-D: 0.404mM, SpTRPA1-DL: 0.116mM, SfTRPA1-DL: 0.321mM). In both isoforms, *S. pallida* TRPA1 was the most sensitive. However, these variations in AITC-induced responses did not hold true for heat-induced currents (Fig. 4D): although we were only able to compare raw current sizes, the response was comparable within each isoform except for the slightly smaller currents of SpTRPA1-CL (Fig. 4D), indicating that evolution of TRPA1 channel property is specific to chemical sensitivity.

On top of the two major L-isoforms, we also investigated the characteristics of two other TRPA1 variants of *S. flava*, SfTRPA1-CS and -DS (hereafter CS and DS). Interestingly, both heat- and chemical-induced current sizes from these two S isoforms were substantially smaller than those of L-isoforms, even at the highest AITC concentrations (Fig. 4E, heat data not shown). Although their activation was weaker, CS and DS exhibited a minimal degree of dose-responses that resembles their counterparts in L-isoforms (Fig. 4E). Overall, our electrophysiological recordings highlighted interspecific and intraspecific variation in functionalities of TRPA1 isoforms that were previously unknown.

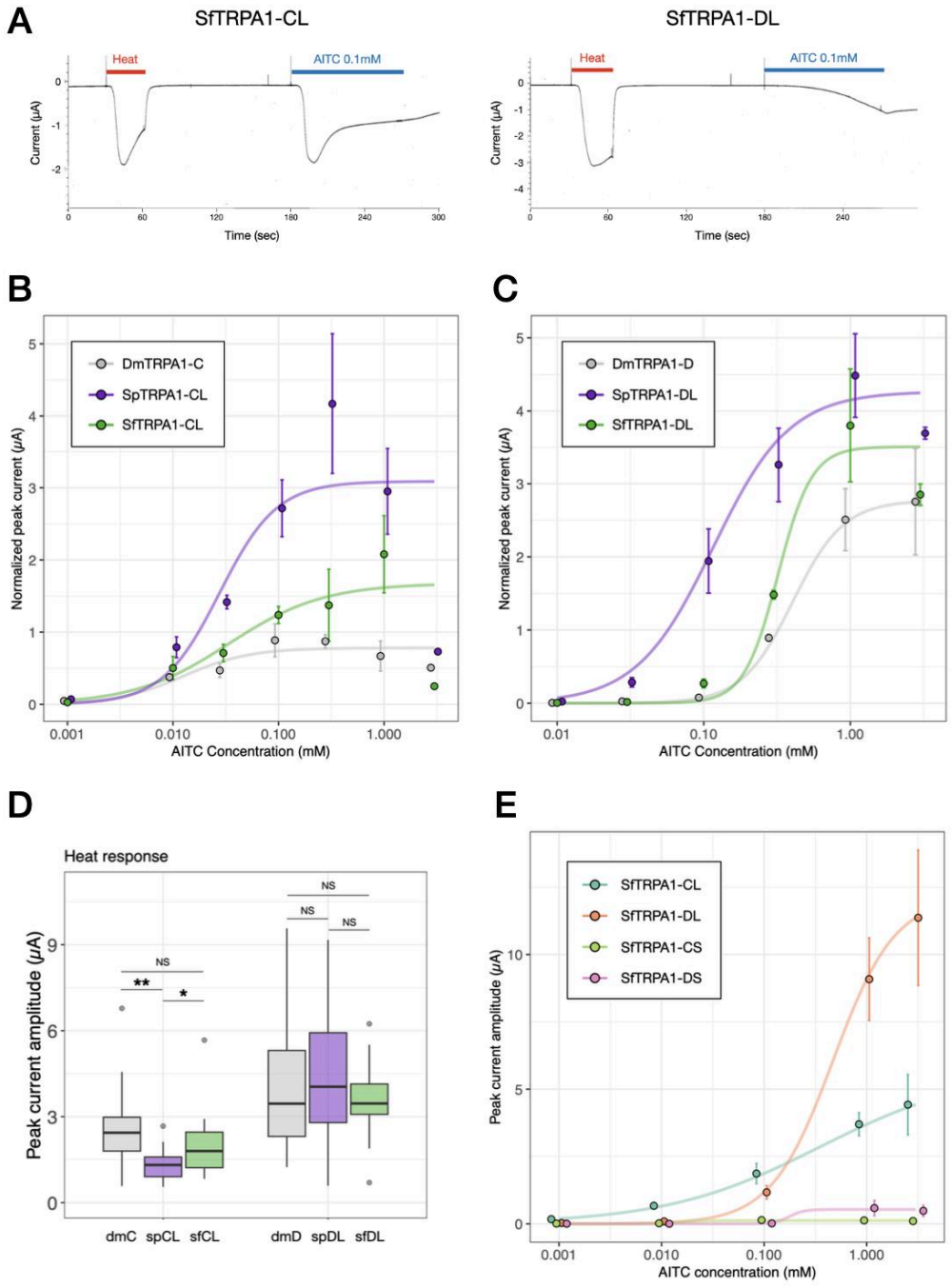


Figure 4. Physiological properties of *Scaptomyza* TRPA1 channels. (A) Example current traces of SfTRPA1-CL and -DL. Each oocyte was initially stimulated by heat, and later by AITC. (B, C) Dose-response curves against AITC of (B) TRPA1-C and (C) TRPA1-D isoforms from the three focal species (n = 3-7 at each concentration, except for 3mM in (B)). Normalized current values were fitted to log-logistic functions with three parameters. (D) Heat-induced raw current sizes of TRPA1-C and -D isoforms. Interspecific differences were analyzed with Kruskal-Wallis test followed by Dunn's test with Benjamini-Hochberg correction. (E) Dose-response curves against AITC of the four SfTRPA1 isoforms, using raw current values.

***In vivo* expression of TRPA1-C isoforms partially recapitulates wild-type behaviors.** Given the functional diversity among TRPA1 isoforms, we next asked whether and to what extent those physiological differences in the ion channel were sufficient to drive phenotypic divergence between species. To address this question, we took a gain-of-function approach using the *GAL4/UAS* system in *D. melanogaster*. Briefly, we first created three homozygous *UAS-TrpA1* lines each carrying a full-length copy of *SfTrpA1-CL*, *SpTrpA1-CL*, or *DmTrpA1-C*, respectively. We then crossed each *UAS-TrpA1* line with a *pGr66a-GAL4* line to drive *TrpA1* expression in all bitter-sensing neurons in the offspring. As controls, *UAS-TrpA1* or *pGr66a-GAL4* lines were also crossed to $TrpA1^{1}$ lines so that no transgene expression was induced in the next generation. These offspring were phenotyped using PER assay to evaluate the contributions of TRPA1 channel on their electrophile sensitivities. We used the non-volatile electrophile NMM as a deterrent instead of AITC, because we wanted to minimize interference on olfactory pathways to focus on the effect of TRPA1 expression in gustatory neurons.

Using this approach, we investigated whether and to what extent *in vivo* expression of TRPA1 channel from the three drosophilid species can recapitulate the patterns we observed from PER or oocyte recordings on wild-type lines. We found that the highest concentration of NMM (10mM) was consistently aversive to all genotypes including *TrpA1*-null control lines, possibly via non-sensory mechanisms like tissue damage (Fig. 5A-D). Still, all TRPA1-expressing lines showed a larger degree of feeding suppression compared to their *UAS*-only controls, confirming the additive effect of transgene expression (Table 1). Preplexingly, we found inconsistencies among the control genotypes in the aversion at 5mM NMM, where $TrpA1^{1}$ and *UAS-DmA1C/+* controls significantly avoided feeding while other controls did not clearly show signs of aversion (Fig. 5A-D). Genotyping using PCR confirmed that all these lines harbor the same *TrpA1*-mutant allele (i.e., $TrpA1^{1}$) as homozygotes (data not shown), which conflicted with previous reports (Kang et al. 2010). In any case, our data suggested that some of our flies retained sensitivity to electrophiles via unknown mechanisms.

Despite this inconsistency, we still discovered some consistent patterns too: The expression of SpTRPA1-CL in *Gr66a*-positive neurons conferred aversion to 5mM NMM, while SfTRPA1-CL expression did not, compared to their respective controls (Fig. 5C, 5D). These findings were aligned with our earlier behavioral and physiological experiments. DmTRPA1-C also did not confer significant aversion at 5mM, possibly due to the aversion seen in *UAS-TrpA1* background alone (Fig. 5B). When comparing the degree of PER reduction between *TrpA1*-expressing and control lines for each TRPA1 channel, SpTRPA1 had the largest effect of transgene at 5mM NMM (Table 1; Δreduction = 0.23-0.065 = 0.165), while SfTRPA1 added up no aversion (Table 2; Δreduction = 0.068-0.07 = -0.02). These values were more comparable at 10mM (Table 2). Taken together, our transgenic PER assay indicated that the physiology of TRPA1

was consistent *in vivo* at least between *S. flava* and *S. pallida*, whereas it does not fully explain the high electrophile aversion of *D. melanogaster*.

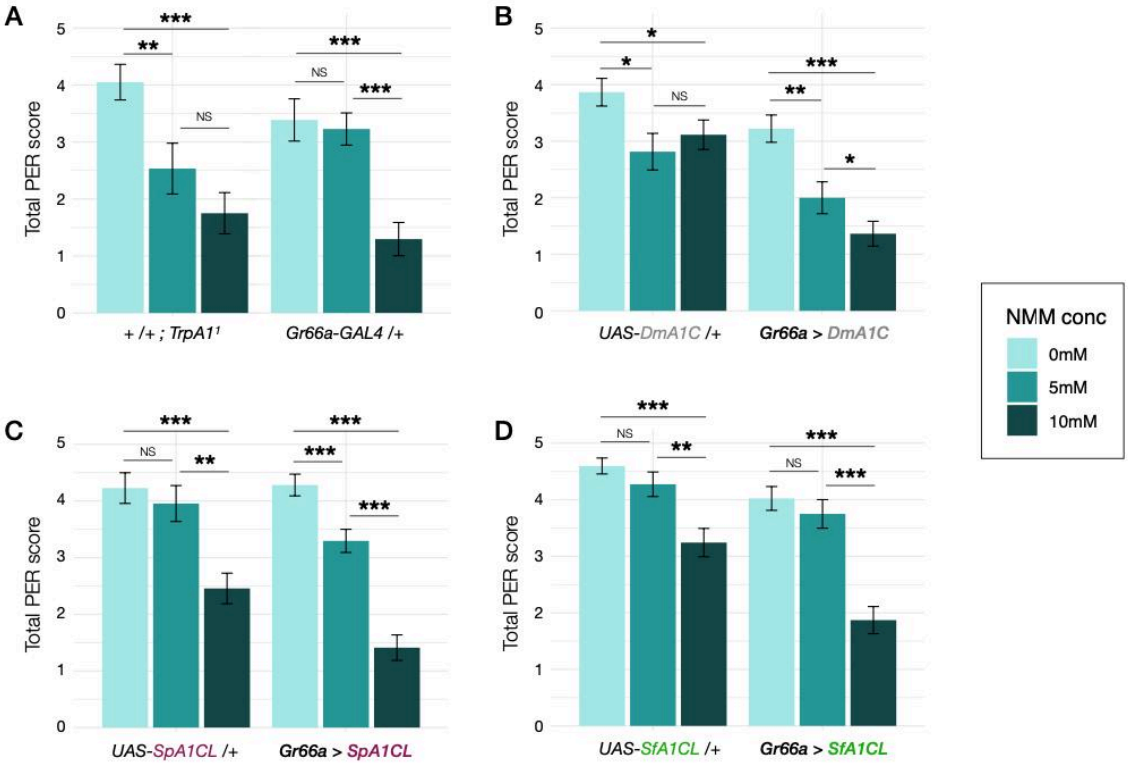


Figure 5. Electrophile sensitivities of *TrpA1*-expressing transgenic flies. (A-D) PER responses against a dosage series of NMM (mixed in 100mM sucrose solution) were tested by pairwise Wilcoxon rank-sum test within each group (P-value: NS, not significant; *, p<0.05; **, p<0.01; ***, p<0.001); (A) Two control genotypes, *TrpA1*^{1} and *pGr66a-GAL4/+*. (B-D) Flies carrying (B) *UAS-DmTrpA1-C*, (C) *UAS-SpTrpA1-CL*, or (D) *UAS-SfTrpA1-CL* fragments with or without *pGr66a-GAL4* cassette to drive their expression in bitter-sensing gustatory neurons. Note that all genotypes carry the same mutant *dTrpA1* allele (i.e., *TrpA1*^{1}) so we can evaluate the contribution of heterologously expressed *TrpA1* isoforms.

Genotype	Average PER score			%Reduction, 0 vs. 5mM	ΔReduction (*GAL4* vs. ctl)	%Reduction, 0 vs. 10mM	ΔReduction (*GAL4* vs. ctl)
	0mM NMM	**5mM NMM**	**10mM NMM**				
+/+;TrpA1[1]	4.05	2.53	1.75	0.374		0.568	
Gr66a-GAL4/+	3.39	3.23	1.30	0.047		0.617	
UAS-DmA1C/+	3.87	2.81	3.11	0.272	0.107	0.195	0.382
Gr66a>DmA1C	3.22	2	1.36	0.379		0.577	
UAS-SpA1CL/+	4.23	3.95	2.46	0.065	0.165	0.419	0.251
Gr66a>SpA1CL	4.28	3.30	1.41	0.23		0.67	
UAS-SfA1CL/+	4.60	4.27	3.24	0.07	-0.002	0.295	0.24
Gr66a>SfA1CL	4.02	3.75	1.87	0.068		0.535	

Table 2. Summary of PER assay in transgenic flies. Rates of PER reduction (%Reduction) were calculated by dividing the average PER score of 0mM by that of 5mM for each genotype. ΔReduction was simply the difference in %Reduction between the test and control lines, supposedly showing the degree of electrophilic aversion added by TRPA1 transgene.

Discussion

Niche shifts in animals are often accompanied by the evolution of behavioral traits (Duckworth 2009), but the genetic changes underlying such events remain elusive. To fill this knowledge gap, here we investigated the molecular basis of gustatory behavior in an evolutionarily young specialist herbivore, *Scaptomyza flava*, and compared it to closely related microbe-feeding fly species. By focusing on the molecular properties of a single ion channel, TRPA1, a primary transducer of electrophilic compounds such as mustard-derived isothiocyanates (ITCs), we aimed to infer part of the evolutionary processes of the herbivorous fly at the molecular resolution. Specifically, we hypothesized that *S. flava* evolved an attenuated aversion toward ITCs due to particular types of molecular changes of the TRPA1 channel – such as the desensitization to the chemical stimulus, the modified compositions of splice isoforms, and the receptor misexpression in sweet-sensing neurons – and addressed each possibility using various methodologies spanning from genomics to electrophysiology.

Our interspecific analyses indicated that the properties of the *TrpA1* gene and its encoding ion channel were indeed correlated with the mustard-feeding ecology of *S. flava*, but they also suggested a slightly complicated evolutionary history. Key findings in

this chapter are summarized as follows: 1) Gustatory avoidance toward electrophiles (AITC) was weaker in *S. flava* than in the two microbe-feeding species (*D. melanogaster* and *S. pallida*), 2) Similar *TrpA1* gene expression patterns were retained in all lineages, but *S. flava* tended to have lower expression levels, 3) *TrpA1* expression was restricted to bitter-sensing neurons in *S. flava* and *D. melanogaster*, 4) Alternative splicing of *Scaptomyza TrpA1* produced species- and tissue-specific compositions of C and D isoforms, 5) Electrophile sensitivity and the current amplitudes of the TRPA1 channel was highest in *S. pallida* and the lowest in *D. melanogaster*, 6) The TRPA1-C isoform responded to AITC at lower concentrations, while D isoform was only activated at higher concentrations, 7) Ectopic expression of the SpTRPA1-CL channel confers electrophilic aversion, while SfTRPA1-CL did not confer clear phenotypic effects. Although wild-type behaviors were consistent with our initial prediction, these data together hinted at a rather complicated trajectory of molecular evolution of *TrpA1* and how it is associated with the colonization of the mustard host plants. To parse out key evolutionary events, we will next discuss the pairwise differences between *S. flava* and either of *S. pallida* or *D. melanogaster*, our two representative microbe feeders.

Differential electrophile sensitivity between *S. flava* and *S. pallida*. Between the two *Scaptomyza* lineages, the functional evolution of TRPA1 is likely associated with the differences in the levels of electrophile aversion. The weaker sensitivity and smaller current amplitudes by AITC stimulation in SfTRPA1 channel compared to SpTRPA1, observed both *in vitro* (Fig. 4B, 4C) and *in vivo* (Fig. 5C, 5D), indicated that these functional changes likely underlie *S. flava*'s attenuated behavioral aversion for ITCs. This is consistent with a previous report on an AITC-insensitive African mole-rat subspecies having TRPA1 channel with low chemical sensitivity (Eigenbrod et al. 2019). Since similar cases were reported in thermal adaptation, where the activation temperatures of TRP channels were either increased or decreased depending on the species' behavior (Gracheva et al. 2010; Akashi et al. 2018; Saito et al. 2022), it seems the modification of ion channel activity is a typical mechanism of sensory evolution. However, although this functional evolution was likely mediated by amino acid substitutions, we do not know which residues were of particular importance. Unlike the African mole-rat TRPA1 (Eigenbrod et al. 2019), SfTRPA1 retains all key cysteine residues in the cytoplasmic domain required for chemical activation of the channel, suggesting that other domains regulating channel activity, such as ankyrin repeats (Cordero-Morales et al. 2011), might be a main driver of the evolution. Further investigation is warranted to identify the exact substitutions of TRPA1 conferring differential chemical sensitivities.

Furthermore, our data highlights an intriguing possibility that the ratio of TRPA1 isoforms might also be at play. Given the lower chemical sensitivity of TRPA1-D isoform, skewed production of *TrpA1-D* mRNA would decrease the electrophile sensitivity of

TrpA1-positive neurons and thus attenuate behavioral aversion of the flies to electrophiles, which should be favored in the *S. flava* lineage since they have to suppress unnecessary aversion to ITCs while laying eggs or feeding on the brassicales. Sensory adaptation aided by functionally variable channel isoforms was reported in TRPV1 of the vampire bat (Gracheva et al. 2011), but is still largely unexplored in many systems, partially due to the difficulty of quantifying long-read transcripts. As alternative splicing profiles of ion channel genes have been slowly characterized thanks to the advances in long-read sequencing technologies (Clark et al. 2020), we believe that our findings pose one of the first few cases of this rare yet potentially prevalent mechanism of ecological adaptation.

Another potential mechanism behind species-specific electrophile preference is the presence of weakly responding TRPA1 isoforms (TRPA1-CS and -DS) in *S. flava* transcriptomes. Similar to the earlier discussion, a higher ratio of such "inactive" TRPA1 isoforms would result in lower neural or behavioral sensitivity to electrophiles. However, this explanation is unlikely because *S. pallida* neurons also expressed another "inactive" isoform, TRPA1-EL, which does not react to any known TRPA1-activating stimuli *in vitro* or *in vivo* (Gu et al. 2019). Our oocyte recordings with mixed isoforms showed that the combinations of SfTRPA1-CL/-CS and SpTRPA1-CL/-EL both fell in between the two channels used (data not shown), suggesting that the presence of EL would lower the activity of bitter neurons to the same degree as TRPA1-CS/DS isoforms would do to *S. flava*. Although Exon S produces CS/DS isoforms specific to *Scaptomyza* genomes, we do not understand the exact roles of these weak isoforms or why the two species express different isoforms. At the very least, we assume their presence alone is not a major driver of phenotypic divergence between the two *Scaptomyza* lineages.

Differential electrophile sensitivity between *S. flava* and *D. melanogaster*. In contrast to the divergence between *Scaptomyza* species, the differences between *S. flava* and *D. melanogaster* are more nuanced. Despite the clear electrophile aversion at the behavioral level (Fig. 1A), our oocyte electrophysiology results showed that DmTRPA1 is actually less sensitive to AITC than SfTRPA1 (Fig. 4B, 4C). This discrepancy indicates that the property of TRPA1 channel alone cannot explain the behavioral differences.

One possible mechanism is the lack of inactive isoforms of TRPA1 in *D. melanogaster* transcriptomes. Although TRPA1-E is expressed in the labellum of *D. melanogaster* (Gu et al. 2019), our data indicate that its proportion is very low (Fig. 3B), thereby electrophile sensitivity of bitter neurons might not be attenuated as much. Furthermore, our RNA-seq data suggested that overall expression level of *TrpA1* is also higher in *D. melanogaster* than in *S. flava* (Fig. 2A), consistent with our observation that *D. melanogaster* labellum is innervated with a higher number of *TrpA1*-positive neurons (Fig. 2B). Possibly, an increased number of sensory neurons might achieve a higher

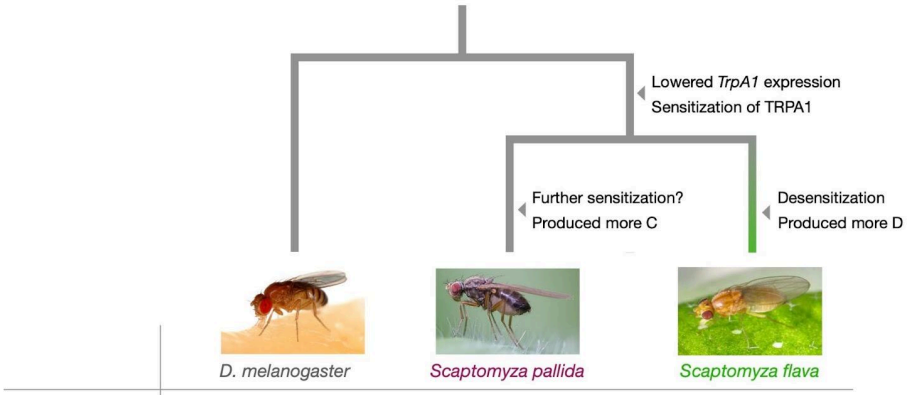
behavioral sensitivity by increasing the frequency of signals sent to secondary neurons or the central nervous system. However, the relationship between sensory neuron numbers and behavioral output might not be so straightforward. A recent study showed that an increase in one type of olfactory sensory neuron in the noni fruit specialist *D. sechellia* (relative to *D. simulans* or *D. melanogaster* lineages) does not increase behavioral sensitivity or projection neuron activity to noni-derived odorants, but instead inhibits the neural adaptation of projection neurons such that flies can track the odorant for a longer period (Takagi et al. 2023). Olfactory and gustatory systems of insects have different neural hierarchies (Su et al. 2009; Scott 2018), thus it is still difficult to predict the consequences of increased neural numbers and a more complete understanding of this will need more experimental studies.

Inconsistency in transgenic PER. Previous studies reported that knockout of endogenous *TrpA1* gene entirely disrupted gustatory aversion against AITC or other electrophilic chemicals in PER assay (Kang et al. 2010; Kohno et al. 2010). Our transgenic assay did not entirely align with these data, since electrophilic aversion was already observed in some of the *TrpA1*-null control flies (Fig. 5). Although our quantification and statistical analysis differed from those of aforementioned studies (i.e., comparing PER responses between 1st trial and 2nd-5th trials), we obtained similar results even after applying these methodologies (data not shown). Why did we end up with these contradictory results? One possibility is the difference in genetic backgrounds of the flies: Although not clearly stated in the manuscripts, authors of these studies likely used *Canton-S* background mutant flies (some information available in (Panzano 2013)), while ours were all backcrossed to w^{1118} line. w^{1118} is commonly treated as wild-type in many studies using transgenic flies, but several studies reported neurobiological and behavioral differences between w^{1118} and other wild-type strains (e.g., (Schilman et al. 2011; Krstic et al. 2013; Qiu et al. 2017)). Since our data suggested that all transgenic flies retained some sensitivity to 10mM NMM even in the absence of TRPA1, this sort of non- TRPA1-mediated aversion might have been augmented in some cases. Further investigation is warranted to gain a fuller picture from our behavioral experiments.

Molecular evolution of *TrpA1* and the colonization to mustard plants. Our comparisons illuminated numerous evolutionary events that occurred in the *TrpA1* gene across the diversification of *Drosophila/Scaptomyza* lineages. We parsimoniously inferred evolutionary timings of these events on the species tree as follows (Fig. 6):
(1) **Ancestral *Scaptomyza* branch**: Reduction of *TrpA1* expression levels and/or neuron numbers, loss of the labellar expression of *TrpA1-A/B* isoforms, gain of Exon S.

(2) **Ancestral *S. pallida* lineage**: Sensitization of TRPA1 channel, skewed expression of TRPA1-C isoform.
(3) **Ancestral *S. flava* lineage**: Desensitization of TRPA1 channel, skewed expression of TRPA1-D isoform.

Considering the putative timing of the colonization to the Brassicales, we propose that the desensitization of TRPA1 and the composition shift to TRPA1-D isoform are two key events aiding the adaptation to ITC-producing plants in the ancestor of the *S. flava* lineage (Fig. 6). Our own data lack information regarding when exactly TRPA1 got sensitized against chemical agonists, but taking into account relatively higher sensitivities in *D. virillis* and *D. mojavensis* TRPA1s than DmTRPA1 (Kang et al. 2010), it is reasonable to infer that TRPA1 became sensitized at the ancestral node of the *virilis/grimshawi/Scaptomyza* clade to electrophiles (presumably to compensate the reduction *TrpA1*-expressing neurons and maintain high aversion to electrophilic toxins), and then desensitized to electrophiles in the *flava* linage to adapt to their mustard-feeding life cycle (Fig. 6). Still, our taxon sampling in this study is minimal, rendering it difficult to track the evolutionary changes of TRPA1 precisely. For example, the mustard-feeding *Scaptomyza* lineage diverged from another herbivorous lineage that specializes in non-ITC-producing plants like *S. graminum*, primarily feeding on the pink family, Caryophyllaceae (Katoh et al. 2017), and it remains an open question whether the molecular evolution of TRPA1 corresponds to the mustard colonization or at the base of the entire herbivorous lineage (as a pre-adaptation to mustards). Future studies will need to cover wider taxonomic ranges to draw a better picture of how a single ion channel gene has been intertwined with dietary evolution of this ecologically diverse fly lineage.



	D. melanogaster	*Scaptomyza pallida*	*Scaptomyza flava*
Aversion	high	high	low
TrpA1 expression	high	medium	low
TRPA1 sensitivity	low	high	medium-low
Isoform composition	N/A	skewed to C	skewed to D

Figure 6. A model scenario of molecular evolution of TRPA1.

Conclusion

The genetic changes underlying the evolution of novel behavioral traits remain elusive. In this chapter, we addressed this question with a mustard-plant specialist fly species, *Scaptomyza flava*, and its two microbe-feeding relatives, *Scaptomyza pallida* and *Drosophila melanogaster*, by comparing the functional properties of the *TrpA1* gene, which encodes a chemosensory ion channel detecting mustard-derived toxins, isothiocyanates (ITCs). Our comprehensive analyses revealed that the overall reduction of chemical sensitivity of TRPA1 channel and the enriched expression of less active TRPA1 isoforms occurred at the lineage leading to *S. flava*, indicating that attenuated aversion to ITCs indeed contributed to the colonization to the mustard plants. However, the distantly-related *D. melanogaster* showed even weaker chemical sensitivity of TRPA1 channel, despite the strong aversion to ITCs at the behavioral level. This discrepancy could be explained by the relatively higher *TrpA1* expression and the increased number of *TrpA1*-positive gustatory neurons in *D. melanogaster*, which might together compensate for the weak activity of the TRPA1 channel. Taken together, our integrative approaches illuminated the multiple molecular mechanisms of the TRPA1 channel's contribution on the gustatory evolution in the mustard-feeding *S. flava*, and highlighted the importance of broad taxon sampling to capture the evolutionary dynamics of a gene function underlying behavioral traits.

Concluding remarks

My dissertation work aimed to uncover the genetic underpinnings of chemosensory evolution behind the dramatic niche shift in the mustard-feeding *Scaptomyza flava*. Based on interspecific comparison between *S. flava* and its microbe-feeding relatives (*S. pallida* or *D. melanogaster*), I took two complementary approaches – transcriptome analyses on the major chemosensory gene families (Chapter 2) and targeted investigations on a receptor gene using multiple experimental tools as a model case (Chapter 3) – to comprehensively capture molecular changes driving the sensory evolution in *S. flava*. These two approaches successfully identified new candidate genes and/or potential molecular mechanisms, such as the receptor desensitization and the modification of isoform composition, which have potentially driven the chemosensory adaptation to the mustard plants. Particularly, Chapter 3 highlighted the importance of integrative experimental approaches to fully capture the molecular evolution behind adaptive traits. In summary, these work together laid out a solid foundation for future functional validations using genome editing or comprehensive receptor deorphanization techniques, which will further dissect the genetic architecture of the behavioral evolution in this charismatic herbivorous species.

References

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, et al. 2000. The Genome Sequence of Drosophila melanogaster. *Science* 287:2185–2195.

Aguilar JM, Gloss AD, Suzuki HC, Verster KI, Singhal M, Hoff J, Grebenok R, Nabity PD, Behmer ST, Whiteman NK. 2024. Insights into the evolution of herbivory from a leaf-mining fly. *Ecosphere* [Internet] 15. Available from: https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecs2.4764

Akashi HD, Saito S, Cádiz Díaz A, Makino T, Tominaga M, Kawata M. 2018. Comparisons of behavioural and TRPA1 heat sensitivities in three sympatric Cuban Anolis lizards. *Mol. Ecol.* 27:2234–2242.

Alexa A. 2006. Gene set enrichment analysis with topGO. *Bioconductor Improv* [Internet]. Available from: https://bioconductor.statistik.tu-dortmund.de/packages/3.3/bioc/vignettes/topGO/inst/doc/topGO.pdf

Almeida FC, Sánchez-Gracia A, Campos JL, Rozas J. 2014. Family size evolution in drosophila chemosensory gene families: A comparative analysis with a critical appraisal of methods. *Genome Biol. Evol.* 6:1669–1682.

Álvarez-Ocaña R, Shahandeh MP, Ray V, Auer TO, Gompel N, Benton R. 2023. Odor-regulated oviposition behavior in an ecological specialist. *Nat. Commun.* 14:3041.

Anders S, Pyl PT, Huber W. 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169.

Anholt RRH, Mackay TFC. 2004. Quantitative genetic analyses of complex behaviours in Drosophila. *Nat. Rev. Genet.* 5:838–849.

Arenas OM, Zaharieva EE, Para A, Vásquez-Doorman C, Petersen CP, Gallio M. 2017. Activation of planarian TRPA1 by reactive oxygen species reveals a conserved mechanism for animal nociception. *Nat. Neurosci.* 20:1686–1693.

Auer TO, Khallaf MA, Silbering AF, Zappia G, Ellis K, Álvarez-Ocaña R, Arguello JR, Hansson BS, Jefferis GSXE, Caron SJC, et al. 2020. Olfactory receptor and circuit evolution promote host specialization. *Nature* 579:402–408.

Auer TO, Shahandeh MP, Benton R. 2021. Drosophila sechellia: A Genetic Model for Behavioral Evolution and Neuroecology. *Annu. Rev. Genet.* [Internet]. Available from: https://doi.org/10.1146/annurev-genet-071719-020719

Baldwin SR, Mohapatra P, Nagalla M, Sindvani R, Amaya D, Dickson HA, Menuz K. 2021. Identification and characterization of CYPs induced in the Drosophila antenna by exposure to a plant odorant. *Sci. Rep.* 11:1–14.

Bendesky A, Bargmann CI. 2011. Genetic contributions to behavioural diversity at the gene–environment interface. *Nat. Rev. Genet.* 12:809–820.

Benton R. 2015. Multigene Family Evolution: Perspectives from Insect Chemoreceptors. *Trends Ecol. Evol.* 30:590–600.

Benton R, Vannice KS, Gomez-Diaz C, Vosshall LB. 2009. Variant ionotropic glutamate receptors as chemosensory receptors in Drosophila. *Cell* 136:149–162.

Bontonou G, Saint-Leandre B, Kafle T, Baticle T, Hassan A, Sánchez-Alcañiz JA, Arguello JR. 2024. Evolution of chemosensory tissues and cells across ecologically diverse Drosophilids. *Nat. Commun.* 15:1047.

Brand P, Hinojosa-Díaz IA, Ayala R, Daigle M, Yurrita Obiols CL, Eltz T, Ramírez SR. 2020. The evolution of sexual signaling is linked to odorant receptor tuning in perfume-collecting orchid bees. *Nat. Commun.* 11:244.

Brown PD, Tokuhisa JG, Reichelt M, Gershenzon J. 2003. Variation of glucosinolate accumulation among different organs and developmental stages of Arabidopsis thaliana. *Phytochemistry* 62:471–481.

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform* 3:lqaa108.

Buck L, Axel R. 1991. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* 65:175–187.

Cameron P, Hiroi M, Ngai J, Scott K. 2010. The molecular basis for water taste in Drosophila. *Nature* 465:91–95.

Carey AF, Wang G, Su C-Y, Zwiebel LJ, Carlson JR. 2010. Odorant reception in the malaria mosquito Anopheles gambiae. *Nature* 464:66–71.

Chaverra-Rodriguez D, Macias VM, Hughes GL, Pujhari S, Suzuki Y, Peterson DR, Kim D, McKeand S, Rasgon JL. 2018. Targeted delivery of CRISPR-Cas9 ribonucleoprotein into arthropod ovaries for heritable germline gene editing. *Nat. Commun.* 9:1–11.

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890.

Clark MB, Wrzesinski T, Garcia AB, Hall NAL, Kleinman JE, Hyde T, Weinberger DR, Harrison PJ, Haerty W, Tunbridge EM. 2020. Long-read sequencing reveals the

complex splicing profile of the psychiatric risk gene CACNA1C in human brain. *Mol. Psychiatry* 25:37–47.

Clyne PJ, Warr CG, Carlson JR. 2000. Candidate Taste Receptors in Drosophila. *Science* 287:1830–1834.

Clyne PJ, Warr CG, Freeman MR, Lessing D, Kim J, Carlson JR. 1999. A Novel Family of Divergent Seven-Transmembrane Proteins. *Neuron* 22:327–338.

Cordero-Morales JF, Gracheva EO, Julius D. 2011. Cytoplasmic ankyrin repeats of transient receptor potential A1 (TRPA1) dictate sensitivity to thermal and chemical stimuli. *Proceedings of the National Academy of Sciences* 108:E1184–E1191.

Croset V, Rytz R, Cummins SF, Budd A, Brawand D, Kaessmann H, Gibson TJ, Benton R. 2010. Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genet.* [Internet] 6. Available from: http://dx.doi.org/10.1371/journal.pgen.1001064

Dahanukar A, Lei Y-T, Kwon JY, Carlson JR. 2007. Two Gr genes underlie sugar reception in Drosophila. *Neuron* 56:503–516.

Diaz F, Allan CW, Matzkin LM. 2018. Positive selection at sites of chemosensory genes is associated with the recent divergence and local ecological adaptation in cactophilic Drosophila. *BMC Evol. Biol.* 18:144.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.

Dobritsa AA, van der Goes van Naters W, Warr CG, Steinbrecht RA, Carlson JR. 2003. Integrating the molecular and cellular basis of odor coding in the Drosophila antenna. *Neuron* 37:827–841.

Doudna JA, Charpentier E. 2014. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346:1258096.

Drosophila 12 Genomes Consortium, Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, et al. 2007. Evolution of genes and genomes on the Drosophila phylogeny. *Nature* 450:203–218.

Duckworth RA. 2009. The role of behavior in evolution: a search for mechanism. *Evol. Ecol.* 23:513–531.

Dudai Y, Jan YN, Byers D, Quinn WG, Benzer S. 1976. dunce, a mutant of Drosophila deficient in learning. *Proc. Natl. Acad. Sci. U. S. A.* 73:1684–1688.

Dweck HK, Ebrahim SA, Khallaf MA, Koenig C, Farhan A, Stieber R, Weißflog J, Svatoš A,

Grosse-Wilde E, Knaden M, et al. 2016. Olfactory channels associated with the Drosophila maxillary palp mediate short- and long-range attraction. *Elife* [Internet] 5. Available from: http://dx.doi.org/10.7554/eLife.14925

Dweck HK, Talross GJ, Wang W, Carlson JR. 2021. Evolutionary shifts in taste coding in the fruit pest Drosophila suzukii. *Elife* [Internet] 10. Available from: http://dx.doi.org/10.7554/eLife.64317

Edger PP, Heidel-Fischer HM, Bekaert M, Rota J, Glöckner G, Platts AE, Heckel DG, Der JP, Wafula EK, Tang M, et al. 2015. The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences* 112:8362–8366.

Eigenbrod O, Debus KY, Reznick J, Bennet NC, Sánchez-Carranza O, Omerbasic D, Hart DW, Barker AJ, Zhong W, Lutermann H, et al. 2019. Rapid molecular evolution of pain insensitivity in multiple African rodents. *Science* 859:852–859.

Ekman P. 2009. Darwin's contributions to our understanding of emotional expressions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364:3449–3451.

Fleischer J, Pregitzer P, Breer H, Krieger J. 2018. Access to the odor world: olfactory receptors and their role for signal transduction in insects. *Cell. Mol. Life Sci.* 75:485–508.

Flint J. 2003. Analysis of quantitative trait loci that influence animal behavior. *J. Neurobiol.* 54:46–77.

de Fouchier A, Walker WB, Montagné N, Steiner C, Binyameen M, Schlyter F, Chertemps T, Maria A, François M-C, Monsempes C, et al. 2017. Functional evolution of Lepidoptera olfactory receptors revealed by deorphanization of a moth repertoire. *Nat. Commun.* 8:15709.

Fujii S, Yavuz A, Slone J, Jagge C, Song X, Amrein H. 2015. Drosophila sugar receptors in sweet taste perception, olfaction, and internal nutrient sensing. *Curr. Biol.* 25:621–627.

Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG. 2008. Drosophila chemoreceptor gene evolution: Selection, specialization and genome size. *Mol. Ecol.* 17:1648–1657.

Garrigan D, Kingan SB, Geneva AJ, Andolfatto P, Clark AG, Thornton KR, Presgraves DC. 2012. Genome sequencing reveals complex speciation in the Drosophila simulans clade. *Genome Res.* 22:1499–1511.

Gloss AD, Vassão DG, Hailey AL, Nelson Dittrich AC, Schramm K, Reichelt M, Rast TJ, Weichsel A, Cravens MG, Gershenzon J, et al. 2014. Evolution in an ancient

detoxification pathway is coupled with a transition to herbivory in the drosophilidae. *Mol. Biol. Evol.* 31:2441–2456.

Goldman-Huertas B, Mitchell RF, Lapoint RT, Faucher CP, Hildebrand JG, Whiteman NK. 2015. Evolution of herbivory in Drosophilidae linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. *Proceedings of the National Academy of Sciences* 112:3026–3031.

Goodwin S, McPherson JD, McCombie WR. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17:333–351.

Grabherr MG;., Brian J. Haas Moran Yassour Joshua Z. Levin Dawn A. Thompson Ido Amit Xian Adiconis Lin Fan Raktima Raychowdhury Qiandong Zeng Zehua Chen Evan Mauceli Nir Hacohen Andreas Gnirke Nicholas Rhind Federica di Palma Bruce W. N, Friedman AAR. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* 29:644–652.

Gracheva EO, Cordero-Morales JF, González-Carcacía JA, Ingolia NT, Manno C, Aranguren CI, Weissman JS, Julius D. 2011. Ganglion-specific splicing of TRPV1 underlies infrared sensation in vampire bats. *Nature* 476:88–92.

Gracheva EO, Ingolia NT, Kelly YM, Cordero-Morales JF, Hollopeter G, Chesler AT, Sánchez EE, Perez JC, Weissman JS, Julius D. 2010. Molecular basis of infrared detection by snakes. *Nature* 464:1006–1011.

Gu P, Gong J, Shang Y, Wang F, Ruppell KT, Ma Z, Sheehan AE, Freeman MR, Xiang Y. 2019. Polymodal Nociception in Drosophila Requires Alternative Splicing of TrpA1. *Curr. Biol.* 29:3961–3973.e6.

Hallem EA, Carlson JR. 2006. Coding of odors by a receptor repertoire. *Cell* 125:143–160.

Harris M, Garcia-Caballero A, Stutts MJ, Firsov D, Rossier BC. 2008. Preferential assembly of epithelial sodium channel (ENaC) subunits in Xenopus oocytes: role of furin-mediated endogenous proteolysis. *J. Biol. Chem.* 283:7455–7463.

Himmel NJ, Letcher JM, Sakurai A, Gray TR, Benson MN, Cox DN. 2019. Drosophila menthol sensitivity and the Precambrian origins of transient receptor potential-dependent chemosensation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* [Internet] 374. Available from: http://dx.doi.org/10.1098/rstb.2019.0369

Hinman A, Chuang H-H, Bautista DM, Julius D. 2006. TRP channel activation by reversible covalent modification. *Proc. Natl. Acad. Sci. U. S. A.* 103:19564–19568.

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS.

Bioinformatics 32:767–769.

Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JMC, Wides R, et al. 2002. The genome sequence of the malaria mosquito Anopheles gambiae. *Science* 298:129–149.

Honeybee Genome Sequencing Consortium. 2006. Insights into social insects from the genome of the honeybee Apis mellifera. *Nature* 443:931–949.

Inagaki HK, Panse KM, Anderson DJ. 2014. Independent, reciprocal neuromodulatory control of sweet and bitter taste sensitivity during starvation in Drosophila. *Neuron* 84:806–820.

Ito H, Fujitani K, Usui K, Shimizu-Nishikawa K, Tanaka S, Yamamoto D. 1996. Sexual orientation in Drosophila is altered by the satori mutation in the sex-determination gene fruitless that encodes a zinc finger protein with a BTB domain. *Proceedings of the National Academy of Sciences* 93:9687–9692.

Jackson FR, Bargiello TA, Yun SH, Young MW. 1986. Product of per locus of Drosophila shares homology with proteoglycans. *Nature* 320:185–188.

Jasti J, Furukawa H, Gonzales EB, Gouaux E. 2007. Structure of acid-sensing ion channel 1 at 1.9 Å resolution and low pH. *Nature* 449:316–323.

Jiang F, Doudna JA. 2017. CRISPR–Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* 46:505–529.

Jiao Y, Moon SJ, Montell C. 2007. A Drosophila gustatory receptor required for the responses to sucrose, glucose, and maltose identified by mRNA tagging. *Proc. Natl. Acad. Sci. U. S. A.* 104:14110–14115.

Jiao Y, Moon SJ, Wang X, Ren Q, Montell C. 2008. Gr64f is required in combination with other gustatory receptors for sugar detection in Drosophila. *Curr. Biol.* 18:1797–1801.

Jordt SE, Bautista DM, Chuang HH, McKemy DD, Zygmunt PM, Högestätt ED, Meng ID, Julius D. 2004. Mustard oils and cannabinoids excite sensory nerve fibres through the TRP channel ANKTM1. *Nature* 427:260–265.

Jourjine N, Hoekstra HE. 2021. Expanding evolutionary neuroscience: insights from comparing variation in behavior. *Neuron* [Internet]. Available from: https://www.sciencedirect.com/science/article/pii/S0896627321000763

Kang K, Panzano VC, Chang EC, Ni L, Dainis AM, Jenkins AM, Regna K, Muskavitch MAT, Garrity PA. 2012. Modulation of TRPA1 thermal sensitivity enables sensory discrimination in Drosophila. *Nature* 481:76–81.

Kang K, Pulver SR, Panzano VC, Chang EC, Griffith LC, Theobald DL, Garrity PA. 2010. Analysis of Drosophila TRPA1 reveals an ancient origin for human chemical nociception. *Nature* 464:597–600.

Karageorgi M, Bräcker LB, Lebreton S, Minervino C, Cavey M, Siju KP, Grunwald Kadow IC, Gompel N, Prud'homme B. 2017. Evolution of Multiple Sensory Systems Drives Novel Egg-Laying Behavior in the Fruit Pest Drosophila suzukii. *Curr. Biol.* 27:847–853.

Kassambara A, Kosinski M, Biecek P, Fabian S. 2017. Package "survminer." *Drawing Survival Curves using "ggplot2"(R package version 03 1)* 3.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30:772–780.

Katoh T, Izumitani HF, Yamashita S, Watada M. 2017. Multiple origins of Hawaiian drosophilids: Phylogeography ofScaptomyzaHardy (Diptera: Drosophilidae). *Entomol. Sci.* 20:33–44.

Khallaf MA, Auer TO, Grabe V, Depetris-Chauvin A, Ammagarahalli B, Zhang D-D, Lavista-Llanos S, Kaftan F, Weißflog J, Matzkin LM, et al. 2020. Mate discrimination among subspecies through a conserved olfactory pathway. *Science Advances* 6:eaba5279.

Khan A, Mathelier A. 2017. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics* 18:287.

Kim BY, Wang JR, Miller DE, Barmina O, Delaney E, Thompson A, Comeault AA, Peede D, D'Agostino ER, Pelaez J, et al. 2021. Highly contiguous assemblies of 101 drosophilid genomes. *Elife* [Internet] 10. Available from: http://dx.doi.org/10.7554/eLife.66405

Kim J, Moriyama EN, Warr CG, Clyne PJ, Carlson JR. 2000. Identification of novel multi-transmembrane proteins from genomic databases using quasi-periodic structural properties. *Bioinformatics* 16:767–775.

Kohno K, Sokabe T, Tominaga M, Kadowaki T. 2010. Honey bee thermal/chemical sensor, AmHsTRPA, reveals neofunctionalization and loss of transient receptor potential channel genes. *Journal of Neuroscience* 30:12219–12229.

Konopka RJ, Benzer S. 1971. Clock mutants of Drosophila melanogaster. *Proc. Natl. Acad. Sci. U. S. A.* 68:2112–2116.

Krstic D, Boll W, Noll M. 2013. Influence of the White locus on the courtship behavior of Drosophila males. *PLoS One* 8:e77904.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921.

Lapoint RT, O'Grady PM, Whiteman NK. 2013. Diversification and dispersal of the Hawaiian Drosophilidae: The evolution of Scaptomyza. *Mol. Phylogenet. Evol.* 69:95–108.

Larsson MC, Domingos AI, Jones WD, Chiappe ME, Amrein H, Vosshall LB. 2004. Or83b encodes a broadly expressed odorant receptor essential for Drosophila olfaction. *Neuron* 43:703–714.

Leung NY, Thakur DP, Gurav AS, Kim SH, Di Pizio A, Niv MY, Montell C. 2020. Functions of Opsins in Drosophila Taste. *Curr. Biol.* 30:1367–1379.e6.

Liao B-Y, Zhang J. 2006. Evolutionary conservation of expression profiles between human and mouse orthologous genes. *Mol. Biol. Evol.* 23:530–540.

Lichtenstein EP, Morgan DG, Mueller CH. 1964. Insecticides in Nature:Naturally Occurring Insecticides in Cruciferous Crops. *J. Agric. Food Chem.* 12:158–161.

Liu L, Leonard AS, Motto DG, Feller MA, Price MP, Johnson WA, Welsh MJ. 2003. Contribution of Drosophila DEG/ENaC genes to salt taste. *Neuron* 39:133–146.

Liu T, Starostina E, Vijayan V, Pikielny CW. 2012. Two Drosophila DEG/ENaC channel subunits have distinct functions in gustatory neurons that activate male courtship. *J. Neurosci.* 32:11879–11889.

Li Y, Hoffmann J, Li Y, Stephano F, Bruchhaus I, Fink C, Roeder T. 2016. Octopamine controls starvation resistance, life span and metabolic traits in Drosophila. *Sci. Rep.* 6:35359.

Li Z, Capoduro R, Bastin-Héline L, Zhang S, Sun D, Lucas P, Dabir-Moghaddam D, François M-C, Liu Y, Wang G, et al. 2023. A tale of two copies: Evolutionary trajectories of moth pheromone receptors. *Proc. Natl. Acad. Sci. U. S. A.* 120:e2221166120.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550.

Lu B, LaMora A, Sun Y, Welsh MJ, Ben-Shahar Y. 2012. Ppk23-dependent chemosensory functions contribute to courtship behavior in drosophila melanogaster. *PLoS Genet.* [Internet] 8. Available from: http://dx.doi.org/10.1371/journal.pgen.1002587

Macpherson LJ, Dubin AE, Evans MJ, Marr F, Schultz PG, Cravatt BF, Patapoutian A. 2007. Noxious compounds activate TRPA1 ion channels through covalent

modification of cysteines. *Nature* 445:541–545.

Makino T, Kawata M. 2012. Habitat variability correlates with duplicate content of drosophila genomes. *Mol. Biol. Evol.* 29:3169–3179.

Matsunaga T, Reisenman CE, Goldman-Huertas B, Brand P, Miao K, Suzuki HC, Verster KI, Ramírez SR, Whiteman NK. 2022. Evolution of Olfactory Receptors Tuned to Mustard Oils in Herbivorous Drosophilidae. *Mol. Biol. Evol.* [Internet] 39. Available from: http://dx.doi.org/10.1093/molbev/msab362

Matsuo T, Sugaya S, Yasukawa J, Aigaki T, Fuyama Y. 2007. Odorant-binding proteins OBP57d and OBP57e affect taste perception and host-plant preference in Drosophila sechellia. *PLoS Biol.* 5:0985–0996.

McBride CS. 2007. Rapid evolution of smell and taste receptor genes during host specialization in Drosophila sechellia. *Proceedings of the National Academy of Sciences* 104:4996–5001.

McBride CS, Arguello JR. 2007. Five Drosophila genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics* 177:1395–1416.

McBride CS, Baier F, Omondi AB, Spitzer SA, Lutomiah J, Sang R, Ignell R, Vosshall LB. 2014. Evolution of mosquito preference for humans linked to an odorant receptor. *Nature* 515:222–227.

McDowell SAT, Stanley M, Gordon MD. 2022. A molecular mechanism for high salt taste in Drosophila. *Curr. Biol.* 32:3070–3081.e5.

Mita K, Kasahara M, Sasaki S, Nagayasu Y, Yamada T, Kanamori H, Namiki N, Kitagawa M, Yamashita H, Yasukochi Y, et al. 2004. The genome sequence of silkworm, Bombyx mori. *DNA Res.* 11:27–35.

Montell C. 2009. A taste of the Drosophila gustatory receptors. *Curr. Opin. Neurobiol.* 19:345–353.

Montell C. 2021. Drosophila sensory receptors—a set of molecular Swiss Army Knives. *Genetics* [Internet] 217. Available from: https://academic.oup.com/genetics/article-abstract/217/1/iyaa011/6128801

Münch D, Galizia CG. 2016. DoOR 2.0--Comprehensive Mapping of Drosophila melanogaster Odorant Responses. *Sci. Rep.* 6:21841.

Ng R, Salem SS, Wu ST, Wu M, Lin HH, Shepherd AK, Joiner WJ, Wang JW, Su CY. 2019. Amplification of Drosophila Olfactory Responses by a DEG/ENaC Channel. *Neuron* 104:947–959.e5.

Obiero GFO, Mireji PO, Nyanjom SRG, Christoffels A, Robertson HM, Masiga DK. 2014. Odorant and gustatory receptors in the tsetse fly Glossina morsitans morsitans. *PLoS Negl. Trop. Dis.* 8:e2663.

Oh SM, Jeong K, Seo JT, Moon SJ. 2021. Multisensory interactions regulate feeding behavior in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* [Internet] 118. Available from: http://dx.doi.org/10.1073/pnas.2004523118

Okamura Y, Dort H, Reichelt M, Tunström K, Wheat CW, Vogel H. 2022. Testing hypotheses of a coevolutionary key innovation reveals a complex suite of traits involved in defusing the mustard oil bomb. *Proc. Natl. Acad. Sci. U. S. A.* 119:e2208447119.

Osborne KA, Robichon A, Burgess E, Butland S, Shaw RA, Coulthard A, Pereira HS, Greenspan RJ, Sokolowski MB. 1997. Natural behavior polymorphism due to a cGMP-dependent protein kinase of Drosophila. *Science* 277:834–836.

Panzano VC (brandeis U. 2013. TRPA1 mediates chemical and temperature sensing in Drosophila.

Peláez JN, Gloss AD, Goldman-Huertas B, Kim B, Lapoint RT, Pimentel-Solorio G, Verster KI, Aguilar JM, Nelson Dittrich AC, Singhal M, et al. 2023. Evolution of chemosensory and detoxification gene families across herbivorous Drosophilidae. *G3* [Internet] 13. Available from: http://dx.doi.org/10.1093/g3journal/jkad133

Peláez JN, Gloss AD, Ray JF, Chaturvedi S, Haji D, Charboneau JLM, Verster KI, Whiteman NK. 2022. Evolution and genomic basis of the plant-penetrating ovipositor: a key morphological trait in herbivorous Drosophilidae. *Proc. Biol. Sci.* 289:20221938.

Pertea G, Pertea M. 2020. GFF Utilities: GffRead and GffCompare. *F1000Res.* [Internet] 9. Available from: http://dx.doi.org/10.12688/f1000research.23297.2

Price PD, Palmer Droguett DH, Taylor JA, Kim DW, Place ES, Rogers TF, Mank JE, Cooney CR, Wright AE. 2022. Detecting signatures of selection on gene expression. *Nat Ecol Evol* 6:1035–1045.

Prieto-Godino LL, Rytz R, Bargeton B, Abuin L, Arguello JR, Peraro MD, Benton R. 2016. Olfactory receptor pseudo-pseudogenes. *Nature* 539:93–97.

Prieto-Godino LL, Rytz R, Cruchet S, Bargeton B, Abuin L, Silbering AF, Ruta V, Dal Peraro M, Benton R. 2017. Evolution of Acid-Sensing Olfactory Circuits in Drosophilids. *Neuron* 93:661–676.e6.

Qiu S, Xiao C, Meldrum Robertson R. 2017. Different age-dependent performance in Drosophila wild-type Canton-S and the white mutant w1118 flies. *Comp. Biochem.*

Physiol. A Mol. Integr. Physiol. 206:17–23.

Richards S, Gibbs RA, Weinstock GM, Brown SJ, Denell R, Beeman RW, Gibbs R, Beeman RW, Brown SJ, Bucher G, et al. 2008. The genome of the model beetle and pest Tribolium castaneum. *Nature* 452:949–955.

Ritz C, Baty F, Streibig JC, Gerhard D. 2015. Dose-Response Analysis Using R. *PLoS One* 10:e0146021.

Robertson HM. 2019. Molecular Evolution of the Major Arthropod Chemoreceptor Gene Families. *Annu. Rev. Entomol.* 64:227–242.

Robertson HM, Warr CG, Carlson JR. 2003. Molecular evolution of the insect chemoreceptor gene superfamily in Drosophila melanogaster. *Proceedings of the National Academy of Sciences* 100:14537–14542.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.

Saito S, Saito CT, Igawa T, Takeda N, Komaki S, Ohta T, Tominaga M. 2022. Evolutionary Tuning of Transient Receptor Potential Ankyrin 1 Underlies the Variation in Heat Avoidance Behaviors among Frog Species Inhabiting Diverse Thermal Niches. *Mol. Biol. Evol.* 39:msac180.

Sánchez-Gracia A, Vieira FG, Rozas J. 2009. Molecular evolution of the major chemosensory gene families in insects. *Heredity* 103:208–216.

Schilman PE, Waters JS, Harrison JF, Lighton JRB. 2011. Effects of temperature on responses to anoxia and oxygen reperfusion in Drosophila melanogaster. *J. Exp. Biol.* 214:1271–1275.

Schmidt HR, Benton R. 2020. Molecular mechanisms of olfactory detection in insects: beyond receptors. *Open Biol.* 10:200252.

Scott K. 2018. Gustatory Processing in Drosophila melanogaster. *Annu. Rev. Entomol.* 63:15–30.

Scott K, Brady R Jr, Cravchik A, Morozov P, Rzhetsky A, Zuker C, Axel R. 2001. A chemosensory gene family encoding candidate gustatory and olfactory receptors in Drosophila. *Cell* 104:661–673.

Shiao M-S, Chang J-M, Fan W-L, Lu M-YJ, Notredame C, Fang S, Kondo R, Li W-H. 2015. Expression Divergence of Chemosensory Genes between *Drosophila sechellia* and Its Sibling Species and Its Implications for Host Shift. *Genome Biol. Evol.* 7:2843–2858.

Shiraiwa T. 2008. Multimodal chemosensory integration through the maxillary palp in Drosophila. *PLoS One* 3:e2191.

Shiraiwa T, Carlson JR. 2007. Proboscis extension response (PER) assay in Drosophila. *J. Vis. Exp.*:193.

Shirai Y, Takahashi M, Ote M, Kanuka H, Daimon T. 2023. DIPA-CRISPR gene editing in the yellow fever mosquito Aedes aegypti (Diptera: Culicidae). *Appl. Entomol. Zool.* 58:273–278.

Starostina E, Liu T, Vijayan V, Zheng Z, Siwicki KK, Pikielny CW. 2012. A Drosophila DEG/ENaC Subunit Functions Specifically in Gustatory Neurons Required for Male Courtship Behavior. *J. Neurosci.* 32:4665–4674.

Su C-Y, Menuz K, Carlson JR. 2009. Olfactory perception: receptors, cells, and circuits. *Cell* 139:45–59.

Sun JS, Xiao S, Carlson JR. 2018. The diverse small proteins called odorant-binding proteins. *Open Biol.* 8:180208.

Takagi S, Abuin L, David Stupski S, Roman Arguello J, Prieto-Godino LL, Stern DL, Cruchet S, Alvarez-Ocana R, Wienecke CFR, van Breugel F, et al. 2023. Sensory neuron population expansion enhances odour tracking through relaxed projection neuron adaptation. *bioRxiv* [Internet]:2023.09.15.556782. Available from: https://www.biorxiv.org/content/10.1101/2023.09.15.556782v1

Talavera K, Startek JB, Alvarez-Collazo J, Boonen B, Alpizar YA, Sanchez A, Naert R, Nilius B. 2020. Mammalian Transient Receptor Potential TRPA1 Channels: From Structure to Disease. *Physiol. Rev.* 100:725–803.

Tarazona S, Furió-Tarí P, Turrà D, Pietro AD, Nueda MJ, Ferrer A, Conesa A. 2015. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res.* 43:e140.

Thistle R, Cameron P, Ghorayshi A, Dennison L, Scott K. 2012. Contact chemoreceptors mediate male-male repulsion and male-female attraction during Drosophila courtship. *Cell* 149:1140–1151.

Tinbergen N. 1963. On aims and methods of Ethology. *Z. Tierpsychol.* 20:410–433.

Toda H, Zhao X, Dickson BJ. 2012. The Drosophila female aphrodisiac pheromone activates ppk23(+) sensory neurons to elicit male courtship behavior. *Cell Rep.* 1:599–607.

Vosshall LB, Amrein H, Morozov PS, Rzhetsky A, Axel R. 1999. A spatial map of olfactory receptor expression in the Drosophila antenna. *Cell* 96:725–736.

Wagner GP, Kin K, Lynch VJ. 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* 131:281–285.

Wang W, Dweck HKM, Talross GJS, Zaidi A, Gendron JM, Carlson JR. 2022. Sugar sensation and mechanosensation in the egg-laying preference shift of Drosophila suzukii. *Elife* 11:e81703.

Wheat CW, Vogel H, Wittstock U, Braby MF, Underwood D, Mitchell-Olds T. 2007. The genetic basis of a plant insect coevolutionary key innovation. *Proceedings of the National Academy of Sciences* 104:20427–20431.

Whiteman NK, Groen SC, Chevasco D, Bear A, Beckwith N, Gregory TR, Denoux C, Mammarella N, Ausubel FM, Pierce NE. 2011. Mining the plant-herbivore interface with a leafmining Drosophila of Arabidopsis. *Mol. Ecol.* 20:995–1014.

Wiens JJ, Lapoint RT, Whiteman NK. 2015. Herbivory increases diversification across insect clades. *Nat. Commun.* 6:8370.

Wittstock U, Agerbirk N, Stauber EJ, Olsen CE, Hippler M, Mitchell-Olds T, Gershenzon J, Vogel H. 2004. Successful herbivore attack due to metabolic diversion of a plant chemical defense. *Proc. Natl. Acad. Sci. U. S. A.* 101:4859–4864.

Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, Bar-Even A, Horn-Saban S, Safran M, Domany E, et al. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* 21:650–659.

Yang K, Huang L-Q, Ning C, Wang C-Z. 2018. Correction: Two single-point mutations shift the ligand selectivity of a pheromone receptor between two closely related moth species. *Elife* 7:1–21.

Yapici N, Cohn R, Schusterreiter C, Ruta V, Vosshall LB, Yapici N, Cohn R, Schusterreiter C, Ruta V, Vosshall LB. 2016. Taste Circuit that Regulates Ingestion by Integrating Food and Hunger Signals Article A Taste Circuit that Regulates Ingestion by Integrating Food and Hunger Signals. *Cell* 165:715–729.

Zhong L, Bellemer A, Yan H, Honjo K, Robertson J, Hwang RY, Pitt GS, Tracey WD. 2012. Thermosensory and Nonthermosensory Isoforms of Drosophila melanogaster TRPA1 Reveal Heat-Sensor Domains of a ThermoTRP Channel. *Cell Rep.* 1:43–55.

Zhou Y, Zhu J, Tong T, Wang J. 2019. A statistical normalization method and differential expression analysis for RNA-seq data between different species. *Biomed. Chromatogr.* [Internet]. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-2745-1